Multi-scale integrated cellular modelling for the study of urban change phenomena

Nuno Eduardo Norte Pinto

MULTI-SCALE INTEGRATED CELLULAR MODELLING FOR THE STUDY OF URBAN CHANGE PHENOMENA

Nuno Eduardo Norte Pinto

December 2015

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola de Doctorat

Curs acadèmic: 2015-2016

Acta de qualificació de tesi doctoral

Nom i cognoms
NUNO EDUARDO NORTE PINTO

Programa de doctorat
GESTIÓN Y VALORACIÓN URBANA Y ARQUITECTÓNICA

Unitat estructural responsable del programa
DEPARTAMENTO DE TECNOLOGÍA DE LA ARQUITECTURA (TA)

Resolució del Tribunal

Reunit el Tribunal designat a l'efecte, el doctorand / la doctoranda exposa el tema de la seva tesi doctoral titulada **Multi-scale integrated cellular modelling for the study of urban change phenomena**

Acabada la lectura i després de donar resposta a les qüestions formulades pels membres titulars del tribunal, aquest atorga la qualificació:

☐ NO APTE ☐ APROVAT ☐ NOTABLE ☐ EXCEL·LENT

(Nom, cognoms i signatura) President/a	(Nom, cognoms i signatura) Secretari/ària	
(Nom, cognoms i signatura) Vocal	(Nom, cognoms i signatura) Vocal	(Nom, cognoms i signatura) Vocal

______________________, _______ d'/de __________________ de _______________

El resultat de l'escrutini dels vots emesos pels membres titulars del tribunal, efectuat per l'Escola de Doctorat, a instància de la Comissió de Doctorat de la UPC, atorga la MENCIÓ CUM LAUDE:

☐ SÍ ☐ NO

(Nom, cognoms i signatura) President de la Comissió Permanent de l'Escola de Doctorat	(Nom, cognoms i signatura) Secretària de la Comissió Permanent de l'Escola de Doctorat

Barcelona, _______ d'/de ____________________ de 2016

Candidate **Nuno Eduardo Norte Pinto**, Lecturer in Urban Planning and Urban Design

Department of Planning and Environmental Management and Centre for Urban Policy Studies, School of Environment, Education and Development, The University of Manchester

Supervisors **Professor Josep Roca Cladera**, Professor of the Department of Architectural Technology

School of Architecture of Barcelona and Centre for Land Policy and Valuation, Universitat Politècnica de Catalunya, Spain

Professor António Pais Antunes, Professor of Spatial Planning and Transport Engineering

Department of Civil Engineering, Faculty of Sciences and Technology, University of Coimbra, Portugal

This doctoral research was done with the valued financial support of the Portuguese *Fundação para a Ciência e Tecnologia* (Foundation for Science and Technology), grant SFRH/BD/37465/2007.

Abstract

The development of urban models based on mathematical and physics concepts has been one of the most intense scientific areas of research for the last two decades in urban studies.

Cellular automata (CA), a mathematical approach to the evolution of systems, is one of these concepts that have gained the attention of geographers and many other urban studies scholars since the 1970s. CA have two main features that are quite interesting for urban modelling. First, CA formulation and early development are very close to the development of computation sciences themselves. Second, CA benefits from an inherent spatiality that suits the modelling of a wide range of spatial phenomena. They allow the simulation of complex patterns of, for example, land use, starting from a very simple and perceivable conceptual framework that includes five simple concepts: (1) the cell and the cell space (representing form); (2) a finite set of cell states (representing, for example, land uses); (3) a neighbourhood of cells (representing spatial interaction); (4) a finite set of transition rules (representing behaviours, the urban function); and (5) the evolution of a system over time (representing the dynamic nature of complex systems).

CA models are commonly used to simulate land use change at a regional or metropolitan level considering land use dynamics at a local level. They consider increasingly smaller cells, making use of the high resolution of today's remote sense images to capture many interactions that occur at a very large scale. Regular cells are used at the local scale (traditionally image pixels) and at a regional scale, as aggregations of smaller cells.

Neighbourhoods are user defined in the majority of the cases, fixing beforehand one of the most important abilities of CA models in capturing spatial interaction and its extent. Transition rules are usually applied to the entire region, making no real difference in the types of interactions that occur at different scales. CA models usually consider external drivers such as accessibility or land suitability as external attributes of cells, disregarding interdependencies between those drivers.

This dissertation presents the research on these previous features by developing a multiscale CA model to simulate land use change both at the regional and at the local scales, taking accessibility not as an exogenous cell attribute but as a part of the modelling package, improving the models' capacity to capture the interdependences between all drivers. The research addressed the issues of scale, cell form, neighbourhood definition, and calibration. A multiscale CA modelling framework aims to simulate land use dynamics at two different spatial and time scales: a macroscale CA that tries to model the aggregated land use change at a regional level; and a microscale CA that tries to model land use allocation at local scale. Irregular cells are used at both scales. Neighbourhood extension is defined at both scales as a model parameter, thus defined by the calibration procedure. The macroscale model generates aggregate values of land use demand as an input for the microscale model, which tries to allocate land use to best fit simulation to reality. Model calibration is made using an optimization procedure based on the particle swarm optimization heuristic.

The dissertation presents and discusses the main features of the models and of the calibration process. A set of modular modelling tools were developed to simulate complex urban phenomena that constitute the foundation of urban growth/urban change. The models have been applied to case studies in Portugal and Spain, with different scales and spatial structures, to illustrate the main findings.

Resumen

El desarrollo de modelos urbanos basados en conceptos matemáticos y físicos ha sido una de las áreas más intensas de investigación científica en el campo de los estudios urbanos. Desde la década de 1970, los autómatas celulares (CA), un enfoque matemático para la evolución de sistemas, es uno de esos conceptos que ha ido ganando la atención de geógrafos y otros académicos dedicados a los estudios urbanos. Los CA tiene dos características bastante interesantes para la modelación urbana. Primero, la formulación y el temprano desarrollo de los CA estuvieron íntimamente vinculados al desarrollo de las ciencias. Segundo, los CA se benefician de una inherente espacialidad adecuada al modelaje de un amplio rango de fenómenos espaciales. Lo anterior permite la simulación de complejos patrones con un marco conceptual simples y perceptible que incluye cinco conceptos simples: (1) la célula y el espacio de la célula (que representa la forma); (2) un conjunto finito de estados de célula (que representa por ejemplo el uso de suelo); (3) una vecindad de células (representando la interacción espacial); (4) un conjunto finito de reglas de transición (representando comportamientos, es decir, la función urbana); y (5) la evolución del sistema a través del tiempo (representando la naturaleza dinámica de sistemas complejos). Los modelos CA se usan comúnmente para simular el cambio en el uso del suelo a nivel regional o metropolitano considerando las dinámicas en el uso del suelo a nivel local. Estos consideran cada vez más, células regulares menores (tradicionalmente imágenes basadas en pixeles) y a escala regional, como agregados de células menores. Las vecindades son definidas en la mayoría de casos por los usuarios,

fijando de antemano una de las habilidades más importantes de los modelos CA para capturar las interacciones espaciales y su extensión. Las reglas de transición por lo general son aplicadas a regiones enteras, sin que haga ninguna diferencia real en el tipo de interacciones que ocurre a distintas escalas. Los modelos CA usualmente consideran los *drivers* externos como la accesibilidad o la idoneidad del suelo como atributos externos de las células, ignorando las interdependencias entre estos *drivers*.

Esta tesis presenta la investigación de las características anteriores, desarrollando un modelo CA de escala múltiple que simula el cambio en el uso del suelo tanto a escalas regionales como locales, tomando la accesibilidad no como un atributo exógeno de la célula, sino como parte de un paquete de modelaje, que mejora la capacidad de los modelos para capturar la interdependencia entre todos los *drivers*. La investigación aborda los temas de escala, la forma de célula, la definición de vecindad, y la calibración. Un marco para el modelaje CA de escala múltiple tiene como objetivo simular la dinámica del uso de suelo a escalas espaciales y temporales diferentes: a nivel macro la CA trata de modelar de manera agregada el uso del suelo, y a nivel macro la CA trata de modelar el uso del suelo a escala local. Células irregulares se usan en ambas escalas. La extensión del vecindad se define a ambas escales como un parámetro del modelo. La escala macro del modelo genera valores de demanda agregada del uso del suelo que son un insumo al modelo de escala micro, que intenta asignar el uso del suelo para que mejor encaje la simulación con la realidad. La calibración del modelo se hace usando el procedimiento de optimización basado en la optimización heurística denominada *particle swarm*.

La tesis presenta y discute las características principales de los modelos y procesos de calibración. Un conjunto de herramientas modulares de modelación fueron desarrolladas para simular fenómenos urbanos complejos que constituyen la base para el cambio y el crecimiento urbano. Para ilustrar los hallazgos más importantes, los modelos fueron

aplicados a estudios de caso en Portugal y España, con diferentes escalas y estructuras espaciales.

Acknowledgements

It is usually acknowledged that Sir Winston Churchill has said "*success is not final, failure is not fatal: it is the courage to continue that counts*", which is considered to be one of Churchill's most famous quotes. Many of Churchill's historians do not credit him for this quote, as there are other records showing that other historical characters have said it. Despite all the controversy, one recognises that this is indeed a motivating quote. And a very illustrative one too.

This quote is, I must say, quite illustrative of the entire process that has been the making of this research and this dissertation and, to some extent, of my career. This incredible and incredibly long hurdle run has been full of great accomplishments and teachings, but had also some misfortunes (and, also, their valuable learnings), as it is, I suppose, the case with the majority of people. We, the Portuguese, call it our *fado*, our fate.

It is when these misfortunes hit one hard that one feels the strength of life itself, the weight of our *fado*, when we need courage to carry on, when we need our loved ones and friends to reach out to help and guide us.

Fortunately, I guess I have been having the courage to carry on during these years despite the big ups and also the big downs that I have experienced, driven by those who love me, by those who I already knew and by those who I have met in the process.

I can only thank you all for this help. But I would also like to say some special words to some.

I start by thanking my main supervisor, Josep Roca, for his friendship and availability since the very first moment we have started exchanging emails, on a Saturday night in February 2006 if I am not mistaken, when I sent him an email showing my interest in doing a PhD at his Centre for Land Policy and Valuation (CPSV), but with conditions, like that I had to have António, my co-supervisor, in this capacity, among other terms. Josep was receptive to my requests and here we are, some years on and a friendship to be proud of. His supervision was fantastic and key to the understanding and completion of this endeavour, his mind-provoking comments and pushes were always insightful and have set in me new standards that definitely shaped my scholarship.

A word of gratitude is due to António Pais Antunes, my co-supervisor. My working relationship with António goes way back and I must say that António has been much more of a friend than just a supervisor or a colleague. I do owe António a great deal of my scientific and scholarly knowledge, skills and capacities, and I can only look forward to continue to learn from and with him about how to be an academic.

Along this process, I got to know new colleagues at CPSV that not only have been very helpful and supportive but also become good friends, always welcoming me while I was in Barcelona. I will not forget the glimpse of a day in the life of my good friend Malcom Burns at Grácia, at the Club Esportiu Europa, and his endless efforts and availability to go with me through the particularities of life; the long dinners fuelled with Latin American and world politics and history with Rolando Biere and Carlos Marmolejo; the nights of glory at Camp Nou with Pau Queraltó (and Josep); the help with data (and patience while helping me!) from Montserrat Moix. What I do know by now is that we will surely keep doing all these things, I can only hope with the same regularity.

I would like to thank some of my former colleagues at the Group of Spatial Planning and Transport Engineering of the University of Coimbra, Ana Bastos, Bruno Santos, Luís Picado, Álvaro Seco, Gonçalo Correia and Arminda Almeida, and at the Department of Civil Engineering of the Polytechnic Institute of Leiria, João Pedro Silva, Anónio Duarte and Luísa Gonçalves for their support and friendship that lasted until today.

A word of gratitude is due to my colleagues at the Department of Planning and Environmental Management of the University of Manchester, where I lecture now. They all have been nothing but supportive and have welcomed me to Manchester since day one with friendship and collegiality. A word is due to Richard Kingston for his mentoring throughout our lecturing walks in Barcelona; to Graham Haughton for his interest and great support as the Head of Department by the time I arrived and for inducting me at Old Trafford football ground; and to Cecilia Wong for her support in inviting and including me as a researcher at the Centre for Urban Policy Studies. To Selma Carson, for her always dedicated interest in my wellbeing in Manchester. To Stephen Hincks and Anna Gilchrist, both guilty of the crime of baking me a birthday cake that is part of the minutes of a department board meeting, my deepest thanking for their wonderful friendship and support during our long nights at Lass O'Gourie. To Alfredo Stein for his friendship, the great discussions we have about life and other interesting issues, when I learn a lot from his incredible knowledge and experience, and for having translated my abstract to Spanish.

Now, nothing of this would make much sense or would have been possible without the love and support of those who are closest to me.

I must thank Helena Nina for her endless support, a loving and unconditional support to me and to my family, which I feel is stronger today than when I started this long journey many years ago.

To Diana, my daughter, I want to say to you how much I love you, for the girl you are, for your unconditional love and incredible patience, a quality rare to find in such a young girl, and that your love is one of the pillars that make me stand and my courage thrive.

To Inês, my dear beloved wife and my best friend, I will only say: we won't give up, we both have the courage to continue.

Contents

List of Figures

List of Tables

1 Introduction

The use of simulation techniques to model complex urban phenomena was, during the last two decades, and still is currently one of the most important areas of research in urban studies. The combination of computer evolution with the introduction to urban studies of sophisticated mathematical techniques – such as cellular automata, from now on referred to as CA, the main concept developed in this dissertation – made possible the development of high resolution models aimed to capture, understand, and forecast the evolution of complex urban phenomena.

The use of microsimulation techniques in urban modelling and simulation is very common currently. From theoretical models (Tobler, 1979, Couclelis, 1985a, White and Engelen, 1993, Batty and Xie, 1997, Benenson and Torrens, 2004) to more or less ambitious operational models that are already being used to support planning processes in many cities around the world (Kakaraparthi and Kockelman, 2011, Koomen et al., 2011, Lavalle et al., 2011), urban simulation is facing the necessity of stepping up to a new level of commitment to planning as a valid tool for supporting decision-making (Waddell, 2001, Friedman et al., 2002, Benenson and Torrens, 2004, Couclelis, 2005, White et al., 2015).

This introduction will focus on the main components of the rationale of the research presented in this dissertation, the concepts of urban modelling and CA. Section 1.1 is dedicated to introducing the broad theme of modelling urban phenomena throughout time. Section 1.2 is a brief introduction to the concept of CA models and to frame their use for addressing the issue of urban growth. Section 1.3 presents an overview of the research process that led to the presentation of this dissertation. Finally, Section 1.4 presents a detailed outline of the dissertation.

1.1 Introducing urban modelling

The use of mathematical tools to model a wide range of spatial problems has been classified for the last decades as an important approach to scientific planning (Batty, 1994). The growth of urban areas is one of the issues which have concentrated a large research effort, with many different approaches being introduced to the field.

CA, the modelling concept that is the centre of this research, is an example of how the field evolved from an initial theoretical perspective towards a more robust application stage, with important research done since at least the 1970s (Tobler, 1970, Tobler, 1979, Couclelis, 1985a, Clarke et al., 1997, Couclelis, 1997a, Batty, 2013, White et al., 2015).

Before the democratization of the use of computers, back in the 1970s, modelling approaches were usually applied on problems which demanded few information and, consequently, small computational capabilities. The information was taken in a very aggregate form, reducing the sensitiveness of those models, which were mainly static and deterministic. The models basically assumed an initial state of equilibrium, both in time and space, evolving to a new equilibrium after an exogenous stimulus. In spite of this technological simplicity, large amount of resources were applied in the development of many models that were, in some cases, being used to simulate real-world cases, especially in North American cities. Nevertheless, the interesting scientific outcomes of that research were not robust enough to convince the policy making and planning communities of their potential as methodological tools to assist their practices. In the mid-1970s, a period of reflection started after the publication of Douglass Lee paper "*Requiem for large-scale models*" (Lee, 1973), where the role of models as tools for scientific planning was severely criticized.

After this first period of large scale modelling that somehow ended by the time Lee published his criticisms, a series of new attempts were made and a new era of modelling began with the new computational resources provided by micro computation. Researchers sought new mathematical approaches founded in much more complex theories, such as discrete choice, agent-based simulation, or cellular automata, as the capacity of processing data at lower costs increased exponentially. At the same time, the development of new computers capable of producing better graphical representations of the problems brought new enthusiasm to the use of models (Klosterman, 1994). The new mathematical theories and modelling tools made feasible the use of disaggregate information and the application of stochastic approaches, reducing the scale of the problems down to agents themselves (individuals, households, trips), thus improving the feasibility of the models (Waddell and Ulfarsson, 2004). It was the beginning of the microsimulation era.

The last three decades were the affirmation era of urban modelling. The consolidation of GIS as the most widespread tool of spatial analysis made possible a new dissemination of the benefits of spatial analysis as a reliable toolkit in planning.

Urban modelling has conquered its place as a part of the current science of planning and cities by now. Many authors have been recently focusing their theoretical proposals using modelling as a major toolkit to address urban issues (Portugali et al., 2012, Batty, 2013, Silva et al., 2014).

Having said that, there is still an important issue that is taking a great deal of attention from researchers regarding the gap between theory (meaning modellers) and practice (meaning planners), which is often considered a major weakness of the use of simulation in planning (Couclelis, 2005). To tackle this gap, models must meet the needs of planning, being able to incorporate values and policy testing in order to enhance their ability to properly simulate complex urban phenomena (Wagner, 1997, Friedman et al., 2002). This evolution in modelling is of great importance for the acceptance and incorporation of robust simulation tools in common planning processes as recent research shows (Brömmelstroet and Schrijnen, 2010, Straatemeier et al., 2010, Brömmelstroet, 2012).

1.2 Introducing cellular automata modelling

CA were first introduced in the 1940s by two mathematicians, John von Neumann, the founder of game theory, and Stanislaw Ulam, who made intensive research in the field of Monte Carlo simulation. Both researchers dedicated part of their scientific work to studying (independently) self-reproduction and to modelling biological life, both trying to devise a mathematical formulation that could reduce the forces governing reproduction to logical rules (Torrens, 2000). The concept of automatic computation is on the very foundations of CA, as von Neumann was strongly influenced by Alan Turing’s Universal Machine, a theoretical concept of a machine capable of universal computing.

> "*A system may be regarded as a universal computer if, given a suitable initial program, it is capable of implementing any finite algorithm through its evolution over time, i.e., that it is capable of producing a working copy as complicated as itself, and the means to make further copies*" (Torrens, 2000: 13).

This machine would be able of taking new actions in a given moment of time based both on a set of parameters and on its knowledge about the external world obtained from an input system of any kind. An automaton can be considered as a Turing-like machine that operates over a structured cell space. The term automaton refers to a self-operating machine that

> "*processes information, proceeding logically, inexorably performing its next action after applying data received from outside itself in light of instructions programmed within itself*" (Levy, 1992 *apud* Torrens, 2000: 15).

However, there is a large difference between CA and Turing's machine: CA are parallel processors, with a series of rules working at the same time while Turing machines are serial processors that can only handle a process after another (Torrens, 2000).

The very idea of having cells that are the ground base of system evolution conceals the notion of space, a concept that played the role of the necessary bridge to the study of geographically-based phenomena. The concept of CA was firstly introduced to geography and urban studies by Waldo Tobler in his seminal work "Cellular geography" (1979). He identified a series of geographical models that were aimed at simulating different spatial phenomena, pointing out one that is the pure application of the CA concept as the best fitted model to cope with the issue of spatial interaction. Tobler's proposal comes after his statement of what is commonly known in quantitative geography as the first law of geography:

> "*everything is related with everything else but closer things are more related than distant things*" (Tobler, 1970: 236).

This understanding of the power of distance in the interaction of spatial features is closely linked with the mathematical concept of CA, as the interaction neighbourhood is key to their implementation.

CA has captured the attention of geographers and planners in the mid-1980s to early 1990s, interestingly at the same time GIS was exploding as the preferred spatial analysis tool in different contexts, from research to all levels of public administration. The seminal work of Couclelis (Couclelis, 1985a, 1987), White and Engelen (White and Engelen, 1993), Batty and Xie (Batty and Xie, 1994) and later Clarke et al. (Clarke et al., 1996) paved the way to the consideration of CA as a powerful tool to model and simulate spatial phenomena of various types. The research on CA gained a new impulse during the 2000s with many researchers exploring further the concept, in a burst that is contemporary of a second wave of fastest and cheapest computational capacities. CA has now countless applications that span from theoretical models (Vancheri et al., 2008, Caruso et al., 2009) to practical applications where they have been used to different extents to support decision making, of which the work developed at the Flemish Institute of Technology, VITO, (VITO, 2015) and at the Research Institute for Knowledge Systems (RIKS, 2015) are very good examples. A large number of publications on CA can be found on a quick search in the Web of Knowledge (by September 2015), with more than 2200 records of CA and all the relevant keywords in urban studies published after the year 2000.

The use of CA provides a high degree of resolution and ensures a high level of representativeness for spatial problems. However, several issues still need careful research, justifying the focus on CA that this research has. The use of irregular cells is poorly treated in the current literature; neighbourhoods are still considered by their mathematical concept rather than by their urban meaning; problem scale is still focused on large metropolitan/regional problems, discarding the consideration of small-scale urban areas; and the use of cellular automata in multi-scale problems – that is, problems where urban development issues are considered simultaneously at regional and at local scale – is rare (Ward et al., 2003).

1.3 The rationale of the research

The combined interest in urban modelling and in the development of quantitative approaches to urban planning, for which CA are a natural concept, are the two pillars of the main aim of this research.

1.3.1 Motivation

Three main issues motivated the definition of this doctoral research project on urban modelling using CA models.

First, the use of simulation techniques applied to urban change/growth problems. The combination of the candidate's background training in Civil Engineering and Urban Planning with his research interests in urban planning led to the consideration of a research subject that could focus on the use of quantitative approaches to urban studies. Urban modelling is currently a mix of theoretical quantitative approaches (with its origins in theoretical geography and in different areas of mathematics, such as operations research) and social-economic sciences, integrating different scientific disciplines, from planning and engineering to economics. However, and despite of the immense diversity of the knowledge in the field, there is still a significant gap between the research and development of urban modelling and its application in common planning practice. The integration of these two areas has become one of the most important issues in the field.

Second, the use of CA still has significant subjects that need further research. CA are usually applied to large regional or metropolitan areas considering large territories and spatial interactions at local scales throughout those larger areas. There is still little work done on multi-scale approaches using CA. The use of local scale CA models is believed to produce good simulation results when considering its integration with other specific models for simulating accessibility, physical suitabilities, or land use demand. The integration of local scale CA with a macroscopic model of urban growth is believed to be an important evolution towards model representativeness. In a regional/metropolitan context it is fair to assume that urban growth depends on the interactions between population and employment at an aggregate level rather than on the interactions between land uses at a local scale. The combination of a macroscale model aimed to analyse and forecast the global trends of urban evolution with a microscale model aimed to assign land uses considering their spatial interaction is believed to be a promising approach for urban modelling.

The use of irregular cells is also a topical line of research. The use of irregular cells based on spatial census tracts is rare among common CA applications and is facing a new impulse with some robust models being applied (Stevens and Dragicevic, 2007, Stevens et al., 2007, Moreno et al., 2008, Moreno et al., 2009, Vliet et al., 2009, Dahal and Chow, 2015). The option for irregular cells that better represent urban structures is aimed to enhance CA abilities of capturing urban change phenomena from real world urban dynamics, increasing its representativeness, especially for important components of CA such as neighbourhood definition or land use interactions.

A set of modular tools specifically designed for simulating transportation systems and assessing accessibility and land use demand are considered in order to enhance model representativeness and to simplify the calibration of the CA models. Calibration is another important issue for modelling urban phenomena. The high degree of interdependence observed between urban phenomena suggests the use of efficient calibration procedures, capable of producing a good search of the highly complex space of solutions.

Third, and final, the possibility of developing theoretical and applied research on urban phenomena and simulation is a challenging topic. There is a strong possibility of combining theoretical research and model development in urban modelling within common planning processes. The incorporation of policy testing and stakeholders' values is another important research interest. As it was mentioned before, this issue is of particular pertinence in urban modelling and simulation. It is important that, despite the simplicity that lies beneath the concept of model development, models could be able to attend to the needs of planning, in particular giving planners the ability to forecast possible scenarios and plausible evolutions towards desirable futures.

1.3.2 Research goals

The main goal of this research is to develop an integrated land use model aimed to simulate urban change phenomena with a multi-scale planning approach. The research is focused on the use of CA as a feasible mathematical concept to model urban change. The main aims of the research are:

- To study the development and application of CA models to the simulation of land use change, explicitly considering accessibility;

- To develop CA models that maintain all the features of the original mathematical concept, without conceptual concessions that could relax their formulations;
- To innovate on the definition and formulation of some of the components of the CA concept thanking into consideration the previous aim;
- To create robust and yet simple to understand land use modelling tools based on the inherent simplicity of the CA concept.

These aims are further detailed and operationalised by a series of research objectives that focus on each component of the research:

- To gain a comprehensive knowledge on the current state-of-the-art on urban modelling as a feasible toolkit in urban studies in general;
- To gain a comprehensive knowledge on the current state-of-the-art on the use of CA models in urban modelling;
- To identify the most challenging gaps in the research on the use of CA models from the knowledge amassed from the two previous objectives;
- To develop further the research on some of the components of the CA modelling concept, namely scale, the definition of cells, the definition of neighbourhood, the formulation of transition rules and the use of calibration;
- To explicitly consider accessibility as an endogenous variable of the CA model, creating a fully-fledged land use transport interaction model using the CA concept;
- To develop and test different CA models that aim to simulate better different contexts of land use change, considering the different spatial interactions that occur at different scales with different degrees of representativeness;
- To develop and apply a microscale CA model aimed at simulating land use change at the local scale;

- To develop and apply a macroscale CA model aimed at simulating the regional/metropolitan spatial interactions of the main drivers of land use change;
- To develop an integrated, multiscale CA-based modelling tool that simulates spatial phenomena at the multiple scales in which each of the phenomenon occur to increase model representativeness

1.3.3 The research process

This research was granted with a doctoral research scholarship by the Portuguese *Fundação para a Ciência e Tecnologia*, the Foundation for Science and Technology (FCT-Portugal), grant SFRH/BD/37465/2007.

The research was related with the development of the research project "*El Proceso de Urbanización en la Costa Mediterránea: ¿Hacia un Modelo Insostenible de Ocupación del Suelo? Un análisis retrospectivo (1956-2006) y prospectivo (2006-2026)*"[1] coordinated by the Centre for Land Policy and Valuations of the Technical University of Catalonia and granted by the Spanish Ministry of Education and Science (grant SEJ2006-09630) as well as other projects centred on urban phenomena. The research was also related with the research project "SOTUR – Strategic Options for Integrating Transportation Innovations and Urban Revitalization", funded by the MIT-Portugal programme with its funds granted by FCT-Portugal.

This dissertation is structured on a series of chapters that are self-contained in their content, thus structured as papers, but interdependent, thus having cross references

[1] The English translation for the title of this project is 'The Urbanization Process in the Mediterranean Coast: Towards an Unsustainable Model of Land Use? Retrospective (1956-2006) and a Prospective Analysis (2006-2026)'

amongst them. The chapters cover different subjects and stages of the work, from the literature review on urban modelling and on CA to the development of the different models and their applications. Therefore, some of the chapters were already published as papers or book contributions while the remaining ones are due to be submitted as papers to leading academic journals immediately after the submission of the dissertation.

1.3.4 The methodological approach

This research is focused on the development and testing of modelling tools based on cellular automata models to simulate urban growth. This focus shapes the methodology by bringing the development of the modelling tools to the forefront of the research objectives. The methodology that was developed and applied is, to a great extent, straightforward. The development of modelling tools implies a comprehensive knowledge of the state-of-the-art that draws from an extensive literature review. This literature review will also provide a full understanding of the main topical issues in both urban modelling and CA modelling, while illustrating the current gaps in their theory and subsequent application to concrete case studies. This will inform the definition of the main research aims (presented in the previous sections). The development of the models implies a constant evaluation of different modelling options that are related to data requirements and availability, representativeness, user interaction and model performance. These options are made considering the existent methodologies that are validated by the literature, to which innovation is added in the areas that this research is exploring some of the identified gaps. Finally, models are tested using a series of theoretical instances and real world case studies to evaluate their performances. Case study selection is made based on the simple criteria of availability and pertinence. Case studies were selected considering the existence of reliable datasets that illustrated interesting land use dynamics in the recent past.

The simplicity of this methodological approach makes irrelevant the existence of a full chapter dedicated to it in this dissertation. Aims and objectives of the research were stated previously in this chapter. Relevant methodological options – all linked to modelling options and case study definition – are detailed and discussed where relevant.

1.4 Detailed outline of the dissertation

This dissertation is structured as a series of self-contained chapters that focus on different components of the research. The chapters are ordered to illustrate the research process, from the understanding of the state-of-the-art in the form of critical literature reviews to the presentation of the different models. This order is also illustrative of the hierarchical dependency of the different CA models, considering their modelling assumptions and formulations and also their scales.

The first two chapters are dedicated to presenting the ongoing academic research and some hints about the practice of urban modelling (Chapter 2) and CA and their applications to urban studies (Chapter 3). The remaining chapters focus on different implementations of CA models. Chapter 4 is dedicated to the presentation of the foundations of CA models based on irregular cells and variable neighbourhoods, with a series of tests made using theoretical test instances and a strong focus on calibration. Chapter 5 is dedicated to the development and application of a microscale CA model, with an application to a transport related case study in the municipality of Coimbra, Portugal. Chapter 6 is dedicated to the development and application of a macroscale CA model, with an application to the Metropolitan Area of Barcelona, Spain. Chapter 7 is dedicated to the development of a multiscale CA model based on the two models presented in the previous two chapters, with an application to the Metropolitan Area of Barcelona. Finally, Chapter 8 is dedicated to the critical analysis of the main findings and to the identification of the future areas or research that arise from this dissertation.

2 Modelling and Urban Studies

Urban modelling has gained in the past two decades a new relevance in urban studies as many of the developments that were still under the research arena have been slowly transferred into the practice. The need to consolidate what Batty called scientific planning (Batty, 1994), supported by the incredible evolution that computers and database management have experience in this period led researchers and practitioners to come closer and to explore further the possibilities of this quantitative approaches.

This chapter is dedicated to an introduction to the topic of the use of modelling in urban studies as a broad concept. The chapter will firstly address the evolution of modelling and the planning process in Section 2.1. In Section 2.2 a brief discussion between the development of models and its integration in the planning process will be addressed. Section 2.3 is dedicated to the use of higher resolution modelling approaches in planning.

Several modelling tools used in simulation will be addressed, such as agent-based and multi-agent simulation, cellular automata, discrete-choice, rule-based simulation, and geographical information system (GIS) based simulation. Finally, some concluding remarks are presented in Section 2.4.

2.1 Modelling and the planning process

From the very early days of planning, like in several other areas of knowledge, a discussion over how it should be approached took place, creating the usual and most certainly necessary tension between theorists and practitioners.

In the specific case of planning, this issue becomes even more complex as it encompasses all sectors of society, from citizens to politicians, from bureaucrats to general stakeholders. Therefore, a necessity for creating a strong and organized planning system emerged from both planning theory and planning practice. As one of the most complex human activities, planning can use all the help it can get from a wide spread areas of knowledge (Couclelis, 2005).

After the blueprint planning era, where the final plan, the final image of the landscape or of the city represented the ultimate goal to achieve, planning become more comprehensive, incorporating social sciences, operations research, economic theories and regional science, trend introduced by the work of Rexford Tugwell and Harvey Perloff at the Chicago School (Klosterman, 1994).

The increasing interest in rational planning (a basic assumption of the Chicago School), was founded on the assumption that scientific approaches and systematic decision-making were the best way to deal with problems in such fields as management, politics and economics. Modelling was considered a major achievement of this scientific approach, therefore a new and important way to assist planning activities (Batty, 1994, Lee, 1994). These "new tools of planning" as Britton Harris called them in 1965, were thought to be a major technological breakthrough that would revolutionize the practice of urban policy making (Wegener, 1994).The idea of planning as a straight and rigid process that lead from a problem to a solution (materialized by the plan), without perturbations of any kind (judgment errors, public participation, monitoring, re-orientation of goals) changed to the cyclic approach proposed for example by McLoughlin (1969).

The introduction of computer sciences in the early 1950s brought new capabilities to mathematical calculus and data processing, limited only by the speed constraints of computer processors and available memory space. By that period the first models of cities were introduced. These models usually focused on transportation and land use allocation problems (Klosterman, 1994). During the 1960s, an important number of cities in the United States already had ambitious land use and transportation models running, and some of these models were already assisting planning activities (Lee, 1973, Batty, 1994, Klosterman, 1994).

In 1973 Douglass Lee published in the Journal of American Planners one of the most important papers ever published in the field, the famous "*Requiem for large-scale models*" (Lee, 1973). He identifies in this paper seven major "sins" of large-scale models: (1) *hypercomprehensiveness*: the early models tried to replicate too large and too complex a system in only one model, in a time when urban knowledge still was taking its first steps; (2) *grossness*: aiming for a large number of results obtained from the models, the information outputted by them was too rough to be use in practice; (3) *hungriness*: these models demand an enormous amount of data; (4) *wrongheadedness*: the models often deduced behaviours for some relationships that could not bet generalized for a different subset of data obtained from the same problem; (5) *complicatedness*: the results of those models were somewhat so complex that they usually need some kind of exogenous intervention that would rebalance the output, with the consequent loss of scientific validity; (6) *mechanicalness*: the systematic errors due to mathematical processes usually produced large amounts of untraceable errors; and finally 7) *expensiveness*: the first theoretical and operational models from the 1960s were too much expensive, in the order of a few millions of dollars. Lee also emphasized the fact that, at the time, no model had produced any kind of relevant theory, as well as no model was founded on strong theoretical grounds.

To make some progress in this field at this critical turning point, Lee draws four major conclusions: (1) models should be more intuitive for potential users; (2) models should combine strong theoretical foundations, objective information and judgment, in order to eliminate the empiricism and the abstract, mainly futile, theorizing; (3) planners should start from simple, well defined problems, towards methods aimed to well identified purposes; and finally (4) models should be simple by nature, since complex models had failed to simulate real life.

Some of the criticism made by Lee where limited by its own boundaries: computational technology was still in a very initial phase, which was, by itself, a strong limitation to the development of models (no matter what the chosen scale was), both in available mathematical tools that were able to be implemented and in data processing capability.

But another author, Gary Brewer, published at the same year of "*Requiem*…" results from his work on the organizational limits to the development of large-scale models. He argued that, rather than theory, technology, data availability, or technical expertise, the inherent difficulties in adapting organizations to technology was the cause for the misuse of models (Batty, 1994).

Batty presents three major achievements that took place in the 1970s, despite this disturbance (Batty, 1994): first, a series of modest but steady refinements on the side of practical applications of land use and transportation models; second, and perhaps the most significant one as Batty argues, the introduction to the urban systems theory of the general concept of optimization, linked to several studies on econometrics and market behaviours; and third, the most challenging one, the incorporation of time – the dynamic behaviour – that could only take place after the development of several new mathematical concepts during the 1960s.

The year of 1973 was the first moment of retrospective for the urban modelling science. Many authors argued opposite opinions on these subjects, but the sense that large scale urban modelling had had its days was evident. Lee's criticisms made a huge impact on the planning community, by the time when the modelling practice had acquired a rudimentary organization, and some tradition and the scientific approach applied to planning was making its first "incursions" in Europe (Batty, 1994). It is symptomatic that the number of papers published on large-scale modelling decreased dramatically, and for a long period of time the subject was, despite a hand full of works, practically put aside (Klosterman, 1994). And Lee's "*Requiem*…" still is, in our days, one of the most cited papers in the field of urban modelling.

One of the consequences of the social change of paradigms that occurred in the 1970s with the end of the post-World War II economic boom and with the early signs of weakness shown by the welfare state was a shift on the planning horizon adopted by the majority of practitioners. The object of planning was re-centred on short term goals, on immediate solutions to problems, rather than to accomplish ambitious long-term strategic objectives, a characteristic of large-scale planning as it was proposed by the Chicago School, for example. It was the shift from planning to management (Batty, 1994).

Another crucial concept of planning had passed virtually untouchable during these turbulent times for the science of modelling: the vital need for large amounts of information. In the 1960s, alongside with modelling techniques, the management of large information systems were the cutting edge of scientific planning (Batty, 1994). The advent of sophisticated data base management systems provided powerful tools for planners as they could now process a larger amount of more disaggregate data in increasingly shorter amounts of time.

The next step that took place in the 1980s was the diffusion of the concept of geographical information systems (GIS). Although, for many years until the mid-1990s, these software tools were mainly used for cartography processing and mapping (Batty, 1994), the integration of new, sophisticated, built in modelling with GIS has provided new grounds for the planning activity (Takeyama and Couclelis, 1997, Wagner, 1997, Batty et al., 1999).

At present, new paths are being explored. Computer capabilities are now at a level which planners and modellers in the 1960s would consider almost science fiction. Whereas location and transportation were the key to the first generation of urban models, the dynamics of growth and diffusion phenomena at a fine scale are the subjects of the new generation of micro-scale modelling (Batty, 2004). Object-oriented programming brought new and powerful tools for modelling at a disaggregate micro-scale (Benenson et al., 2002, Barros, 2005, Benenson et al., 2005, Semboloni, 2005). Micro-simulation is now a reality, supported by a series of techniques such as cellular automata and agent-based simulation. GIS and data base management are two of the most developed areas in software, with a series of commercial products that easily and by lower costs provide the necessary data processing capacity. The developments in the last two decades both in data availability and computational capacity have created a big impulse on the use of models, clarifying their role as a scientific approach to the highly comprehensive planning process (Couclelis, 2005).

Models are shifting from a comprehensive perspective, the main assumption of the first generation of models, to a sketch-planning-type modelling, oriented for solving closely adapted local situations, standing out policy-oriented, practical goals rather than broad strategic goals, a characteristic of the former ones (Batty, 2004, Couclelis, 2005).

But some of the critics formulated in the 1970s remain present. There are some concerns that modellers still focus their main attention on model development rather than on the planning problem underlying the model. Urban simulation is as much a planning exercise in simulation as it is in planning sciences (Torrens and O' Sullivan, 2001). Couclelis stated that in all but trivial cases, the hope on good predictive models in the field of social phenomena is lost (Couclelis, 1997a). There still is an apparent paradox on the fact that, as well intentioned modellers were creating land use models aimed to forecast future states of complex systems, these models helped little (or not helped at all) planners in their tasks (Couclelis, 2005).

2.2 The theory/practice dichotomy

The important notion of the conflict that exists in urban modelling between theoretical and practical perspectives, between the use of models and the practice of planning, is very well described by Couclelis:

> "*Models are based on science; planning is about policy. Models are much better (…) at dealing with natural science problems; planning is mired in difficulties most often due to issues in the purview of social sciences. Models are usually developed from within particular disciplinary perspectives; planning must integrate across all domains. Models are about information and facts; planning is about interpretation*

> *and values. (…) Models codify uncertain knowledge; planning must lead to certain action. (…)"* (Couclelis, 2005: 1359)

This natural tension between theorists and practitioners, between modellers and planners was already mentioned in the previous section. This continuous tension between modelling and planning results from those general dichotomies mentioned before between science and policy, natural and social sciences, between analysis and synthesis, studying the past and preparing the future (Couclelis, 2005). Modelling was firstly considered the new grounds for a new scientific approach to planning (Batty, 1994, Wegener, 1994). Despite of the disturbance experienced by urban modelling throughout the past decades (or, in a more realistic perspective, since almost the beginning of its practice), the practice of planning was always considered intimately dependent of its theory, only varying the scale and the extent of dependence the later imposed to the former.

There still is a structural gap between planning theory and practice. It only emphasizes the mistrust in the scientific approach made by the use of models. Modellers and planning support systems (PSS) developers must try hard on developing solutions that meet the planners needs (Couclelis, 2005). Therefore, applicability must be one of the major aims (if not the goal to achieve) of current and future studies in urban modelling, as it constitutes the means that will provide the needed validation to this scientific area. Modellers must accept that the role of scientific planning goes far beyond the implementation of forecasting models; at the same time, planners must find a balance between participation and systematic expertise (Couclelis, 2005).

The shift of paradigm from the pioneers of the 1960s to the current practice is mainly associated with the general perception that rational planning, aiming to understand and control the entire system, failed to accomplish the needs of more modest, incremental interpretation of planning (Wegener, 1994). In this context of integration of stakeholders, a major aim of current planning practice, the recent development of new methodologies for the integration of stakeholders' values, such as multi-criteria evaluation and the new Value Sensitive Design methodology (Friedman et al., 2002) can improve the use of simulation in the planning process (Waddell and Ulfarsson, 2004). It is also very important to capacitate models to evaluate objectives that are stated in planning policies, even when they are somehow undefined (Waddell and Ulfarsson, 2004).

The integration of stakeholders and their value systems, alongside with the fact that planning generates a variable set of goals and actions, imposes the consideration of uncertainty as a key factor for the success of the relationship modelling/planning (Couclelis, 2005). Three major roles for land use models are proposed by Couclelis in order to increase their mission to support planning: scenario writing (what may be), visioning (what should be) and storytelling (what could be) (Couclelis, 2005). Scenario writing is a notion that has its roots on modelling science, being one of its historical goals. Visioning is useful to integrate community interests and values in order to try to reach broad consensus on strategic matters. Storytelling can help to compare future desired and feared evolutions, in realistic terms that could effectively assist the planning process.

Since those uncertain times, a small but steady increase has been registered in the field of urban modelling. Time had tempered the experience and what had been accomplished between 1973 and the early 1990s could not be considered as failure (Batty, 2004). This increase promoted by the spectacular development of computational capacities, was studied by Wegener for the special issue of the Journal of American Planners published in 1994 (Wegener, 1994). Wegener assembled a list of groups and locals that were developing and implementing urban models of any kind with an operational perspective (see Figure 2.1).

The criteria for choosing a particular group was: (1) that a mathematical model implemented on a computer and aimed to analyse past evolution and to forecast future urban scenarios should be the basis of the urban model itself; (2) that the modelling approach should have a comprehensive framework, integrating all the essential processes of urban phenomena; and (3) were excluded all those works that only presented theoretical work without any operational implementation, as valid as those works could be. It is interesting to notice that by that time, 20 years after Lee's "*Requiem*…" an important group of scientists was developing urban modelling and had constituted an informal scientific network, crucial for the strengthening and increasing gain in coherence of this scientific field (Wegener, 1994).

In the present, the number of active groups working on integrated urban models has reached the highest point, with a series of operational models implemented all over the world. Waddell, Bhat, Ruiter, Bekhor, Outwater and Schroer (2001) reports a series of ongoing work on the field: the UrbanSim framework is in operational use in the Puget Sound Region, in the state of Washington (Waddell and Ulfarsson, 2004); the Reusable Modelling Components for Land Use, Transportation, and Land Cover project is dedicated to the development of robust and modular set of modelling tools capable of being replicated in different urban and regional contexts. Miller et al. (2004) refers some integrated land use and transportation software packages that are already available, as MEPLAM (Hunt and Simmonds, 1993) and TRANUS (de la Barra, 2001), although these models are founded on aggregate approaches with strong equilibrium assumptions on several variables of the systems (Miller et al., 2004).

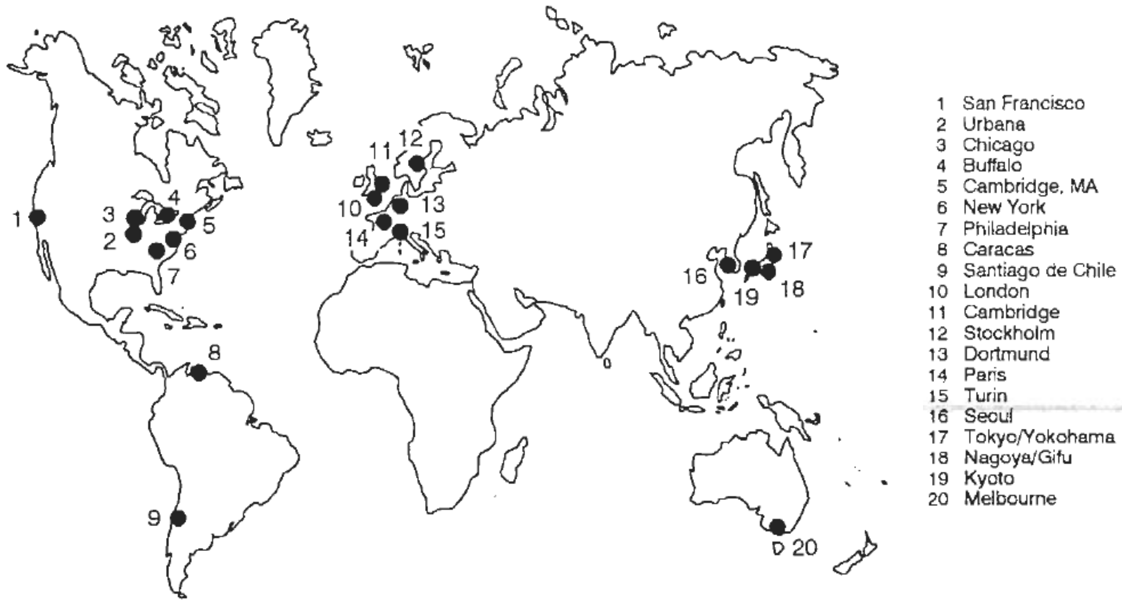


Figure 2.1 Map of active urban modelling centres (Wegener, 1994)

Agarwal, Green, Grove, Evans and Schweik (2002) proposed a classification methodology for land use/land cover change models. They established a series of key factors that influence model performance by analysing several agricultural, forestry, and land use models. Their work was of particular interest because they established three major vectors for the classification of land use models: space complexity, temporal complexity and decision making complexity. Although it seems obvious that these three components are determinant for model classification, it was necessary to establish parameters that could classify a model considering different degrees of complexity. The relationship between these three vectors of complexity can be easily interpreted from the scheme depicted Figure 2.2.

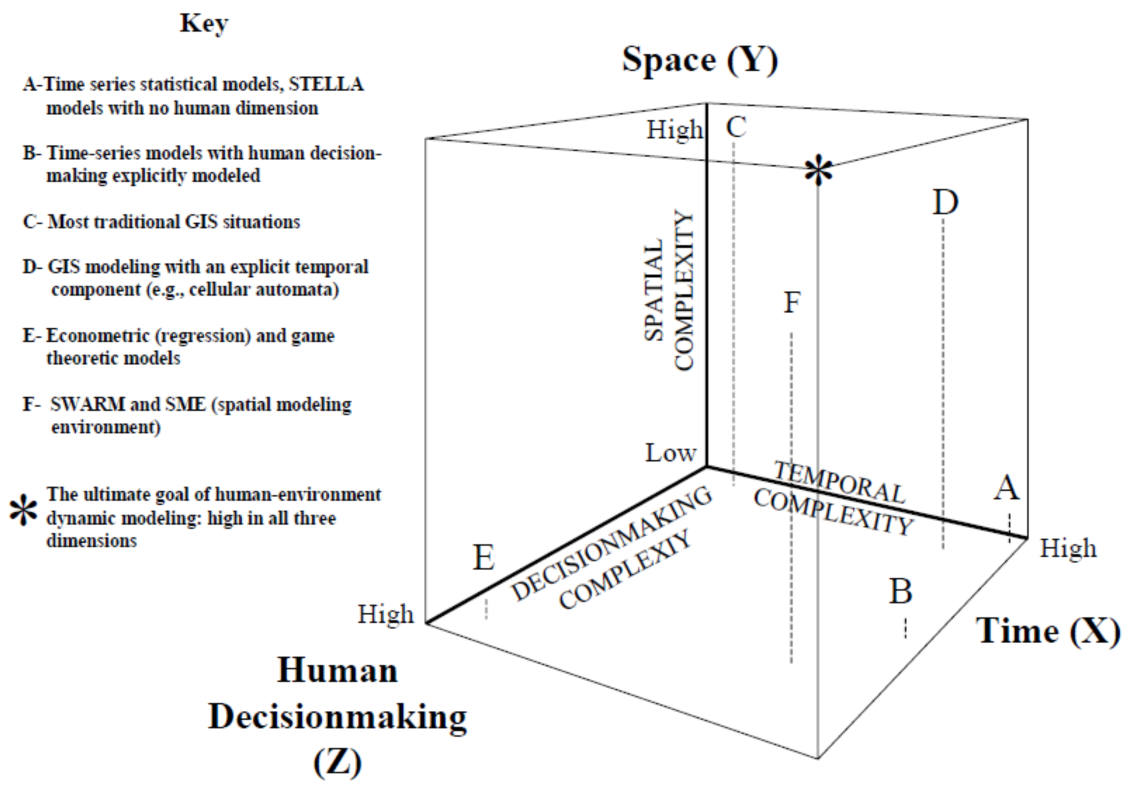


Figure 2.2 Three-dimensional framework for reviewing and assessing land-use change models (Agarwal et al., 2002)

Six degrees of classification for human decision making complexity were proposed: from grade 1 (No human decision making – only biophysical variables in the model) up to grade 6 (Multiple types of agents whose decisions are modelled overtly in regard to choices made about variables that affect other processes and outcomes).

The study analysed nineteen models that cover different modelling approaches. It classified these models and compared them considering a series of human drivers or social patterns and preference variables. An intuitive analysis was presented in the form of comparison graphics where the entire set of models was compared considering not only their spatial and temporal levels of complexity but also a series of other characteristics (some examples are depicted in Figure 2.3).

The study pointed out some orientations for the future in land use modelling. The use of open-source platforms was referred as a proficient path to pursue in order to enhance the collaborative basis of land use modelling (some modelling techniques are currently embracing this paradigm, for example agent-based simulation (Swarm, 2002)). This new vision of modelling could also deal with the concurrent work developed by independent research groups on common modelling techniques. Another important conclusion is related with the incorporation of land use policy in model development. Models should meet the needs of policy makers and planners in order to enhance their applicability for assisting the planning process.

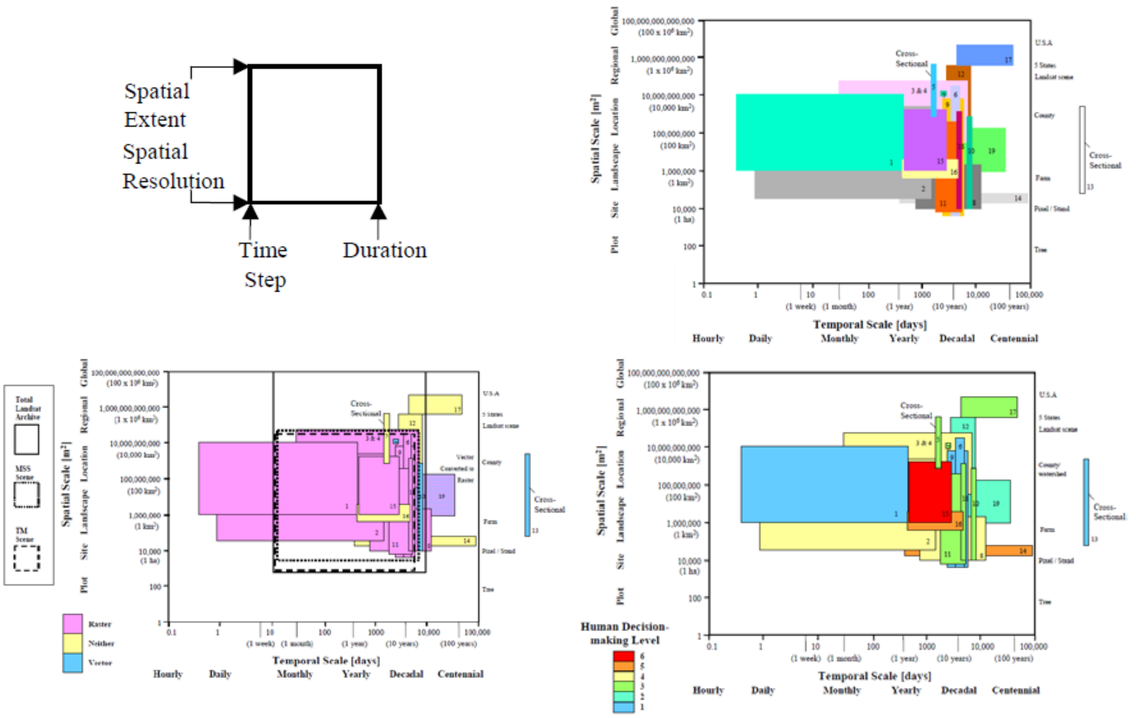


Figure 2.3 Comparative graphic for human decision making complexity (for the complete list of models see Agarwal et al. (2002))

Klosterman and Petit updated the list of significant urban models that are currently being implemented (Klosterman and Petit, 2005). They assembled this list considering simultaneously the modelling approach of each work and the main task purpose by each model, which is summed up in Table 2.1 and Table 2.2.

Complexity is also one of the major investigation areas currently under attention. Many authors identifies complexity as the key factor to understand urban phenomena (White and

Engelen, 1997, Benguigui et al., 2000, Batty and Torrens, 2001, Li and Yeh, 2001a, De Keersmaecker et al., 2003, Batty, 2005b).

Table 2.1 Categorization of selected planning support systems

Technique	Task			
	Land use/Land cover change	Comprehensive projection	3D visualization	Impact assessment
Large-scale urban	METROPILUS SPARTACUS TRANUS UrbanSim	METROPILUS SPARTACUS TRANUS UrbanSim		
Rule-based	CUF WhatIf?™1.1	WhatIf?™ 2.0	CommunityViz	CommunittyViz INDEX© Place^{3}S
State-change	CUF II CURBA			
Cellular automata	SLEUTH DUEM			

Source: Klosterman and Petit (2005)

Cities, like several spatial phenomena, are complex systems. “The mixture of different urban activities creates a logic of its own, but a logic nonetheless. As almost all cities presents this complexity, it is reasonable to suppose that complexity is somehow one of its essential qualities” (White and Engelen, 1993). Complexity expresses itself through spatial scale: from local scale behaviours of individuals (vehicles or people) emerge structured and ordered patterns in aggregate large scale. In fact, the whole concept of complexity hinges on the notion of emergence (Torrens, 2000). One of the goals of complexity studies is to derive universal laws of complex systems from common principles based on simple features (Torrens, 2000). However, many processes in the natural world may not be deduced in universal laws capable of granting a decent theoretical basis of complex systems (Casti, 1997) cited by (Torrens, 2000).

Table 2.2 Information sources for selected planning support systems

Model	References	Web-site
CommunityViz	(Kwartler and Bernar, 2001)	http://www.communityviz.com
CUF, CUF II and CURBA	(Landis, 2001)	
DUEM	(Xie, 1996)	http://lgre.emich.edu
INDEX©	(Allen, 2001)	http://www.crit.com
METROPILUS	(Putman and Shih-Liang, 2001)	
Place[3]S	(Snyder, 2001)	http://www.energy.ca.gov/places
SLEUTH	(Clarke et al., 1997, Silva and Clarke, 2002)	http://ngcia.ucsb.edu/projects/gig
SPARTACUS	(Latuso, 2003)	http://www.fhwa.dot.gov/planning
TRANUS	(de la Barra, 2001)	http://www.modelistica.com
UrbanSim	(Waddell, 2001)	http://www.urbansim.org
WhatIf? ™	(Klosterman, 2001, Klosterman et al., 2003)	http://www.what-if-pss.com

Source: Klosterman and Petit (2005)

The best way to characterize a complex system is identifying the states it can take and the conditions to take them. This can be easily understood with a simple example based on cellular automata. Considering a system with *n* elements (for instance cells) describing a particular state, each state described by a binary existence (say developed) or otherwise for each element, then there are 2^n distinct states. If a lattice of some hundreds cells is considered, with a wider set of rules for transit between those states, then conventional theorizing cannot describe the problem (Batty and Torrens, 2001).

Complex systems include two key elements (Batty and Torrens, 2001): first, "system extensiveness" along any spatial, temporal or topical dimension[2], as complex systems cannot be reduced or aggregated without loss of their structure; second, process, meaning that space and time dynamics suffers unexpected changes, often followed by emergence.

[2] The fact that complex systems are allowed to evolve without the constraints of reductionism runs directly against the traditional scientific paradigm of analysing the essence of phenomena to deduce theory.

2.3 The microsimulation approach

In Section 2.1 the evolution of urban modelling throughout the past decades was shortly described, since the early efforts from the 1950s to the new paths based on powerful and inexpensive computation. This evolution kept up with the evolution of computer science, benefiting with the new calculus resources provided by computer.

The transition from the early large scale modelling phase, based on equilibrium assumptions and deterministic approaches gave place to the fine scale based modelling supported by new scientific approaches. Microsimulation, which was developed in the 1960's, was only applied to urban modelling later in the 1980's. Since then, the development of discrete choice modelling and the emergence of cellular automata and multi-agent simulation techniques have created a proliferation of modelling approaches (Waddell and Ulfarsson, 2004).

Waddell and Ulfarsson (2004) presents a series of preliminary model design choices that must be considered in urban modelling. These assumptions establish the difference between macrosimulation and microsimulation, as they set up a series of orientations that are thought to adjust the simulation to the reality they aim to simulate.

The first choice regards behavioural resolution. Systems can be considered working at an aggregate scale of average behaviour or they can be assumed working at a disaggregate level, based on individual agents. Secondly, the simulation must be based on deterministic or on stochastic behaviour: deterministic models are commonly used along with aggregate scale of behaviour since the average behaviours can be easily approximated with fixed rates of change. Finally, issues related with the resolution of agents, space and time must be pondered. Simulation systems range, in general, from macroscopic resolution to microscopic resolution.

Macroscopic systems have larger units of analysis, dealing with aggregate information both spatially and statistically and they are essentially static and deterministic. The low consumption of data and computational resources make macroscopic simulation one of the most widely used approaches (Waddell and Ulfarsson, 2004).

Microscopic models have, in opposition to the previous scale, small units of analysis. These scale level of modelling is the strongest beneficiary of the evolution of computation throughout the last twenty years. As computational resources evolved and become less expensive and faster, microscopic models started to increase the amount of data processed and to deepen the resolution of the models, with a consequent increase in their feasibility. These models have a stochastic behaviour as they support clearer behavioural specifications (Waddell and Ulfarsson, 2004).

In between these two modelling scales there is an intermediate mesoscopic scale. This term is essentially used to classify models that integrate characteristics of both macroscopic and microscopic models. They can present large analysis units with small time steps or they can use aggregate data for some aspects of the model at the same time they use detailed information for other aspects (Waddell and Ulfarsson, 2004).

Most geographic theories are static where rational actors were assumed to interact in a market that remains in a state of equilibrium. This is not a reasonable way to describe a city, which common sense and experience tell us is rarely if ever in an equilibrium state. Almost all cities are undergoing continual growth, change, decline and restructuring, usually simultaneous (White and Engelen, 1993).

The assumption that urban systems are better represented by dynamic, stochastic, high resolution models along with impressive developments on computation made microsimulation the most fitted approach for dealing with these issues. Applications based on microsimulation are being developed both in theoretical and operational perspectives in different areas of urban sciences. Transport systems analysis produced a series of real-time applications, based on individual agents (see Miller et al. (Miller et al., 2004) for a brief list of models). Integrated land use models are also being developed in the last few years using microsimulation, such as UrbanSim, SLEUTH and WhatIf? operational models (see Table 2.1 on page 28).

Although microsimulation was introduced in the 1960s, it was only in the 1980s that it started to be used in urban modelling (Waddell and Ulfarsson, 2004).

The early modelling approaches were based on techniques as spatial interaction, spatial input/output and linear programming. Spatial interaction is founded on the gravity model applied to model trip destination choices or residential and employment location. These models are limited in the degree of spatial detail used and do not represent the behavioural factors influencing the phenomena they try to simulate, especially market and prices behaviours (Waddell and Ulfarsson, 2004). Spatial input/output models are an extended application of the input/output model of the US economy presented by Leontief to represent spatial patterns of location for economic activities and people and goods movements between zones (Leontief, 1966). This approach has the merit of including explicit real estate and labour markets, as well as travel demand, but it still considers different states of equilibrium for changes in the model inputs (Waddell and Ulfarsson, 2004). Another technique used in early models, although less often, is linear programming. This technique is focused on the optimization of an objective function and it is more suited to the exploration of land use alternatives that optimize some urban function (such as travel cost), than to simulate complex and realistic behavioural responses to input changes (Waddell and Ulfarsson, 2004).

Microsimulation as a modelling approach essentially implies the use of individual-level scales. Waddell and Ulfarsson (2004) describes the most important microsimulation techniques currently under use.

Discrete choice modelling is a standard method whenever the behaviour of individuals (households, people, and trips) is modelled, particularly after the publication of the Random Utility Theory by Daniel McFadden (Waddell and Ulfarsson, 2004). Models such as logit and nested logit are frequently used to predict individual choices among finite sets of alternatives, a very common goal on travel demand and mode choice modelling.

There are several land use models developed in recent years using GIS and a rule-based set of procedures to allocate population, employment, and/or land use. Examples include the CUF model (Landis, 2001) and WhatIf? (Klosterman, 2001). These applications may have a useful role in making models more accessible, but there is a risk that model users would interpret the models as having a more behavioural basis than their rules actually contain (Waddell and Ulfarsson, 2004).

Multi-agent simulation (a generalization of agent-based simulation) is another simulation method available that works on a disaggregate level. It draws upon complex systems theory, focusing the modelling on the emergent systems behaviour arising from the interactions between agents (Waddell and Ulfarsson, 2004). This techniques has been the object of intensive research since the creation of the Swarm software environment for implementation of models of this type (Swarm, 2002). There is extensive ongoing research on these methods with promising results (Barros, 2005, Benenson et al., 2005, Semboloni, 2005).

Cellular automata modelling has emerged from the field of complex systems theory as a means of representing emergent properties derived from sets of simple behavioural rules operating over a cell-based pattern. This approach was introduced to spatial problems by

Waldo Tobler in his ground breaking work "*Cellular geography*" (Tobler, 1979) and has been widely applied since then in many land use and land cover change problems (Couclelis, 1985a, White and Engelen, 1993, Xie, 1996, Clarke et al., 1997, O' Sullivan, 2001b, Barredo et al., 2003, Batty, 2005a).

Cellular automata present many advantages for modelling urban phenomena. Their conceptual simplicity and the high level of spatial resolution can be pointed out as two of their main features. It is a decentralized approach, it provides a link to complexity theory, and it makes the connection of form with function and of pattern with process. It has good visualization characteristics, it is a flexible and dynamic approach and, more important, it is based on a set of very simple elements (Torrens and O' Sullivan, 2001). Cellular automata also have advantage when facing the traditional simulation approaches based on differential or difference equations: it is inherently spatial, with rule-base dynamics, with a much higher computational efficiency which means that dynamics can be modelled with very high spatial resolution (White and Engelen, 1993, Batty et al., 1997). But these advantages yields simultaneously its handicap: cellular automata models are constrained by their own simplicity and their ability to illustrate real world phenomena is often diluted by their abstract characteristics (Torrens and O' Sullivan, 2001).

2.4 Conclusions

This paper aimed to briefly introduce the theme of modelling and urban studies. The goal was to gather key information and literature references on the subject and to allow the readers to establish a starting point for the study of the field.

Key issues were addressed. A brief evolution of the use of models in urban studies was presented and the dichotomy between theory (regarding modelling) and practice (in planning) was discussed. This discussion is considered of great importance as it defines the real extent of the applicability of models in urban studies. The tension observed by many authors between modellers and planning practitioners was (and still is) at the same time the weakness and the strength of urban modelling. The scepticism with which many planners still face the use of models creates a difficult framework to the introduction of modelling approaches. On the other hand, the necessity for proving models' validity and applicability is a major motivation for modellers to carry on in the search for more sophisticated, collaborative, and effective modelling tools.

Finally, the large set of modelling tools that are currently under use gives hope to the future of urban modelling. Different approaches are been carried out in an increasingly integrated fashion. Collaborative approaches are in current development and the effective integration of human decision making and policy issues with spatial and temporal complexity provides models a whole new set of possibilities. Issues regarding spatial and temporal resolution still are in the centre of theoretical discussion. And the goal of developing modular modelling tools capable of being applied to a wide set of different urban contexts is one of the most challenging line of research in the field.

3 Cellular Automata in Urban Simulation: Basic Notions and Recent Developments

3.1 Introduction

The use of mathematical tools to model a wide range of spatial problems has been classified for the last decades as an important approach to scientific planning (Batty, 1994). The growth of urban areas is one of the issues which have concentrated a large research effort, and a series of new modelling techniques were introduced and developed throughout the last decades. Before the democratization of the use of computers, back in the 1970s, modelling approaches were usually applied to problems which demanded few information and, consequently, small computational effort. Computers made possible the use of more disaggregate information, shifting models from a large-scale perspective to microsimulation. New techniques were sought and Tobler (1979) proposed the application of a cellular approach for modelling geographic phenomena, introducing the use of cellular automata to geography. During the next two decades, a great effort was made to develop CA-based models: Couclelis (Couclelis, 1985a), White and Engelen (White and Engelen, 1993), and Batty (2005b) worked on their theoretical issues regarding CA application to urban studies, Batty and Xie (1997) and Clarke *et al.* (1997) worked on the application of important evolutions of CA to real world problems, Semboloni (2000) studied urban infrastructure development, O'Sullivan (2001b) used an integrate approach based on CA and on graph theory to study gentrification, Semboloni (1997) and Ward, Murray and Phinn (2003) developed multi-scale urban models based on CA, and Silva and Clarke (2002) and Barredo et al. (2003) made applications of previously developed CA models to large metropolitan areas.

Cellular automata (CA) were first introduced in the 1940s by John von Neumann, the founder of game theory, and Stanislaw Ulam, who made intensive research in the field of Monte Carlo simulation. Both researchers were dedicated to studying self-reproduction and to modelling biological life, trying to devise a mathematical formulation that could reduce the forces governing reproduction to logical rules (Torrens, 2000). The concept of automatic computation is on the very foundations of CA, as von Neumann was strongly influenced by Alan Turing's Universal Machine, a theoretical concept of a machine capable of universal computing. "*A system may be regarded as a universal computer if, given a suitable initial program, it is capable of implementing any finite algorithm through its evolution over time, i.e., that it is capable of producing a working copy as complicated as itself, and the means to make further copies*" (Torrens, 2000). This machine would be able of taking new actions in a given moment of time based both on a set of parameters and on its knowledge about the external world obtained from an input system of any kind. An automaton can be considered as a Turing-like machine that operates over a tessellated cell space. The term automaton refers to a self-operating machine that "*processes information, proceeding logically, inexorably performing its next action after applying data received from outside itself in light of instructions programmed within itself*" (Levy, 1992, cited by Torrens, 2000). However, there is a large difference between CA and Turing's machine: CA are parallel processors, with a series of rules working at the same time while Turing machines are serial processors that can only handle a process after another (Torrens, 2000).

This chapter is dedicated to introducing the concept of CA and its main applications to quantitative approaches in geography and urban studies. The chapter has two main parts. Sections 3.2 and 3.3 focus on basic notions and on the introduction of CA to geography. The concept and its first applications in geography are discussed in detail, after which some applications are presented. Section 3.4 is dedicated to identifying some trends for future research in CA, pointing out some issues that need to be addressed. Finally, some concluding remarks are presented in Section 3.5.

3.2 The concept of cellular automata

According to the first formulation of a 2D cellular automata, due to von Neumann, each automaton cell processes information and performs action considering both a set of transition rules and data received from the environment. Formally, each cell *A* is defined by a set of cell states $S = \{S_1, S_2, ..., S_N\}$ and a set of transition rules *T*

$$A \leftarrow (S, T) \tag{3.1}$$

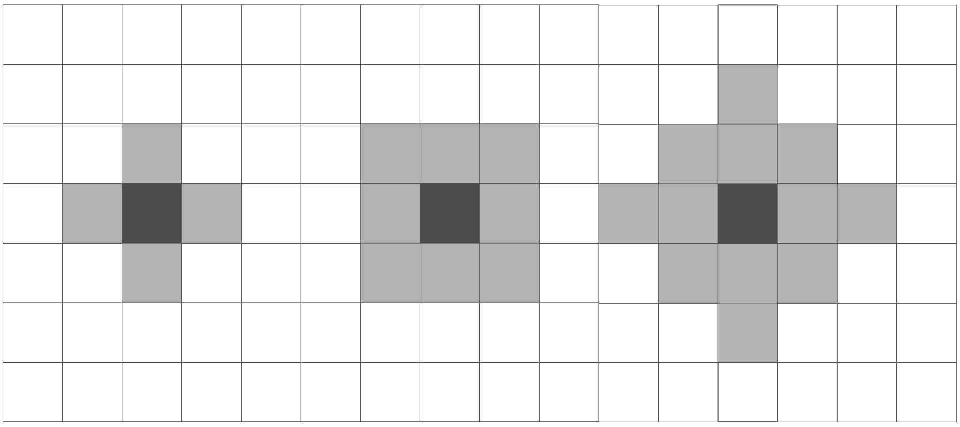
Transition rules define an automaton state S_{t+1} at time step $t+1$ depending on its state at time step t, S_t ($s_t, s_{t+1} \in S$) and an input information I_t:

$$T : (S_t, I_t) \rightarrow S_{t+1} \tag{3.2}$$

A grid of automata becomes cellular automata when the input is influenced by the states of neighbouring cells. This definition of neighbourhood is basically the set of cells that influences the state of the cell under consideration.

$$A \leftarrow (S, T, R) \tag{3.3}$$

where R denotes automata neighbourhood A, and establishes the boundary for drawing the input information I necessary for the application of transition rules T. Classic neighbourhoods such as the von Neumann neighbourhood (Figure 3.1a) and the Moore neighbourhood (Figure 3.1b) are commonly used in 2D cellular automata. The first one is the set of four adjacent cells (usually) corresponding to the main cardinal points. The second one is the entire set of eight adjacent cells. Variations and conjugations of these two types of neighbourhood have also been used (e.g., the one depicted in Figure 3.1c).

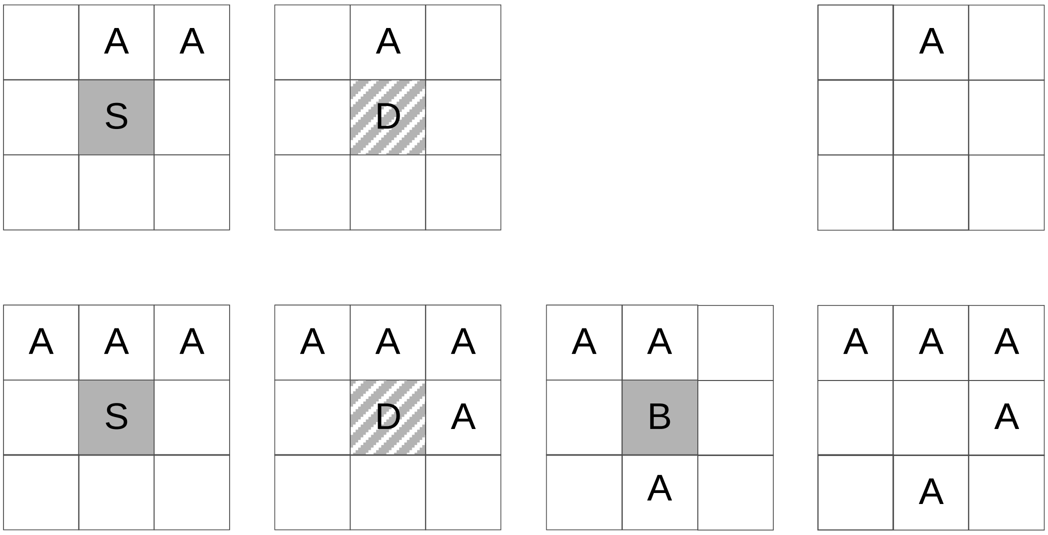


(a) (b) (c)

Figure 3.1 Classic von Neumann's, Moore's and mixed neighbourhoods for 2D CA

There are two main moments in the history of CA after the introduction of complex systems theory: the discussion over John Conway's Game of Life in 1970 and the work of Stephen Wolfram in the 1980s (Wolfram, 1984).

The Game of Life was aimed to design a simple set of rules to study microscopic spatial dynamics of population (Benenson and Torrens, 2004). This 2D CA was based on a set of two cell states, alive (1) or dead (0), and on three plus one transition rules: survival (Figure 3.2a), death (Figure 3.2b) and birth (Figure 3.2c). The cell remained dead if none of these rules were applied (Figure 3.2d). The survival rule states that a living cell survives if it is surrounded by two or three living cells; the death rule states that a cell can die of loneliness (if it only has one neighbour) or because it is overcrowded; the birth rule states that a new cell is born if it has exactly three neighbour living cells; and finally, no cell can be born if it only has one neighbour or if it has more than five neighbours.



(a) Survival **(b) Death** **(c) Birth** **(d) Dead**

Figure 3.2 Conway's Game of Life set of rules

From this simple set of rules, the Game of Life could generate extremely complex patterns of growth through iterative simulation. Furthermore, it could be easily replicated following the concept of the Turing Machine. Its main achievement was to formulate a simple interdisciplinary tool for representing complex spatial systems and for modelling their dynamics (Benenson and Torrens, 2004). Conway pursued a configuration that could generate moving configurations of stable patterns. Conway challenge on the pages of

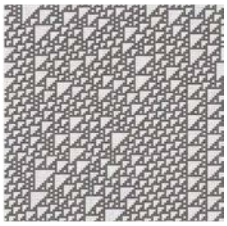
Scientific American for the research of such a configuration was a defining moment on CA history. Robert Gosper and his team at the Massachusetts Institute of Technology implemented a version of Life-CA that was capable of generating a machine that could reproduce copies of itself that were as complicated in their structure (Torrens, 2000). Considering the Game of Life as the first widely known CA, and relating its classification as a game to represent world evolution, it is interesting to consider that, perhaps, CA is like the world in its ability of representing the most complex forms of evolution from simple, well-understood interaction rules (Couclelis, 1997a). It is worth stating the procedures of the Game of Life as they embody the key elements of CA (Batty et al., 1997).

Following the introduction of CA as a suitable tool for studying complex systems dynamics made possible through the discussion over the Game of Life, research on the limits of system's spatial patterns began. The mathematician Stephen Wolfram made extensive research on CA exploring the final configurations towards which they could evolve to from simple sets of parameters and rules. Using a simple 1D CA, where the neighbourhood is the set formed by the cell and its adjacent neighbours (a set of 3 cells), Wolfram studied limiting patterns for 256 different transition rules (Wolfram, 1983, 1984). Based on numeric experiments, Wolfram demonstrated that the final configuration of CA does not depend on the initial state of cells but is defined by transition rules. Wolfram classified CA according to their dynamic behaviour and the patterns generated by them, identifying the four main classes described and depicted in Table 3.1.

Several other studies tried to devise new classifications by extending the study of Wolfram's classes to larger neighbourhoods and different sets of rules and parameters. Wolfram's classification has the disadvantage of making it impossible to classify a given CA based on the transition rule alone, not knowing what the spatial pattern would be

(Benenson and Torrens, 2004). However, Wolfram's classification remains the most popular one.

Table 3.1 Wolfram's classification for 1D CA behaviour

Class	I	II	III	IV
CA dynamics evolves towards	Spatially stable pattern – each cell reaches the stable value of "0" and "1"	Sequence of stable or periodic structures – each cell changes its states according to the fixed finite sequence of "0"s and "1"s	Chaotic aperiodic behavior – the sequence of cell state is not periodic, but the spatial patterns repast themselves in time	Complicated localized structures, which are sometimes long-lived and are more complex than those of Class III
Type of system dynamics	Limit points	Limit cycle	Chaotic attractors	Attractors unspecified
Graphic Output				

It is also important to mention the classification introduced by C. Langton which is very interesting for geography and urban studies (Benenson and Torrens, 2004). He introduces the concept of inactive cells, i.e. cells that cannot change state during CA evolution. Although a series of new classifications were proposed, several studies on higher-dimensional CA showed that they were similar to Wolfram's classification for 1D CA (Benenson and Torrens, 2004).

3.3 Cellular automata and urban studies

Most geographical theories are static, assuming that the interactions of agents take place in a generic market that remains in a state of equilibrium; this assumption is far from being reasonable, as all cities are continually undergoing complex changes. This fact makes imperative the use of dynamic models based on the processes that occur in the territory (White and Engelen, 1993). The great appeal of CA relates to the fact that many classes of system dynamics can be simulated through it. Another important feature of CA is their ability to give equal weight to the importance of space, time, and system attributes (Batty et al., 1997). The natural ability that CA have to represent systems with complex spatial/temporal behaviours from a small set of simple rules and states made this technique very interesting for geographers and urban researchers. CA are intrinsically spatial and they are used to model a wide range of phenomena due to their ability to represent spatial processes, from forest fires to epidemics, from traffic simulation to regional-scale urbanization, polycentricism, gentrification, historical urbanization, urban growth, form and location (White and Engelen, 1993, Torrens and O' Sullivan, 2001). CA-based modelling also allows the integration of socio-economic and natural systems models in a detailed and realistic way (White and Engelen, 1997).

There are three main classes of urban CA models that are a direct result of the exploration of modifications of the conventional CA: (1) models designed to explore spatial complexity; (2) models designed to research themes of economical, sociological and geographical areas; and (3) models designed to produce operational tools for planning (Torrens, 2000). In this section, three pioneer studies are presented, exploring the formulation and the spatial complexity of CA from a theoretical standpoint.

From the 1960s until the mid-1980’s geographers were more focused on the study of comprehensive regional models (Benenson and Torrens, 2004). The criticisms formulated by Douglass Lee (1973) in his “*Requiem…*” established a transition point after which the modelling community started to question the necessity for large scale models and shifted to small scale problems. Problem complexity, even when a small set of spatial relationships was used for modelling purposes, and the lack of experimental data for model calibration generated a search for simpler approaches that could produce more reliable simulations (Klosterman, 1994). CA presented an opportunity to deal with these new modelling requirements. Tobler and Couclelis had set the ground for this cellular approach, in the 1970s and the 1980s, when it really had a major development along with computer graphics, fractal theory, chaos, and complexity.

Waldo Tobler presented in 1970 his study on population growth simulation based on a cellular model (Tobler, 1970). While addressing the interdependence of population growth in a given location and its dependence on the population growth everywhere else in the world, he stated the first law of geography: *everything is related to everything else, but near things are more related than distant things*. This is an important concept that supports the link between CA and geography because of the importance of the concept of spatial neighbourhood. Instead of using trend equations with limited sets of coefficients that could surrogate the real behaviour of growth in every location, he considered the influence on population growth limited to a neighbouring set of locations. Later in 1979, Tobler published another important study where CA are explicitly considered on the study of spatial phenomena (Tobler, 1979). Tobler made a classification of cell models for land use simulation, taking into consideration their dependence on spatial and temporal dynamics (Figure 3.3). Model type I is an independent model, where the land use for a given location $g_{i,j}$ at time $t+\Delta t$ is not related with the situation at time t. Model type II is classified as functionally dependent because land use at a given cell $g_{i,j}$ at time $t+\Delta t$ depends on the land use at that location at time t. Model type III is named a historical model as land use for a given location $g_{i,j}$ at time $t+\Delta t$ depends on a series of previous states at previous time periods. Model type IV is classified as multivariate because it depends on several variables other than land use at that location i,j. Finally, model type V is classified as the geographical model because land use at a given location $g_{i,j}$ at time $t+\Delta t$ is dependent on the land uses in a given neighbourhood of that location at time t.

Tobler developed his study using a model type V which he classified as a dynamic one, because land use at a given location g_{ij} at time step $t+\Delta t$ is a function F of land use at that location at time t and of a measure of the influence of all land uses located at the neighbouring cells, n_{ij}, as expressed in Equation 4:

$$g_{ij}{}^{t+\Delta t} = F(g_{ij}{}^{t}, n_{ij}) \quad (3.4)$$

The neighbourhood n is based on the traditional von Neumann neighbourhood and is defined as a geographical domain of influence. He considers that, because of the different notions of neighbourhood different residents have, it should be possible to have dynamic neighbourhoods varying on size, shape and orientation.

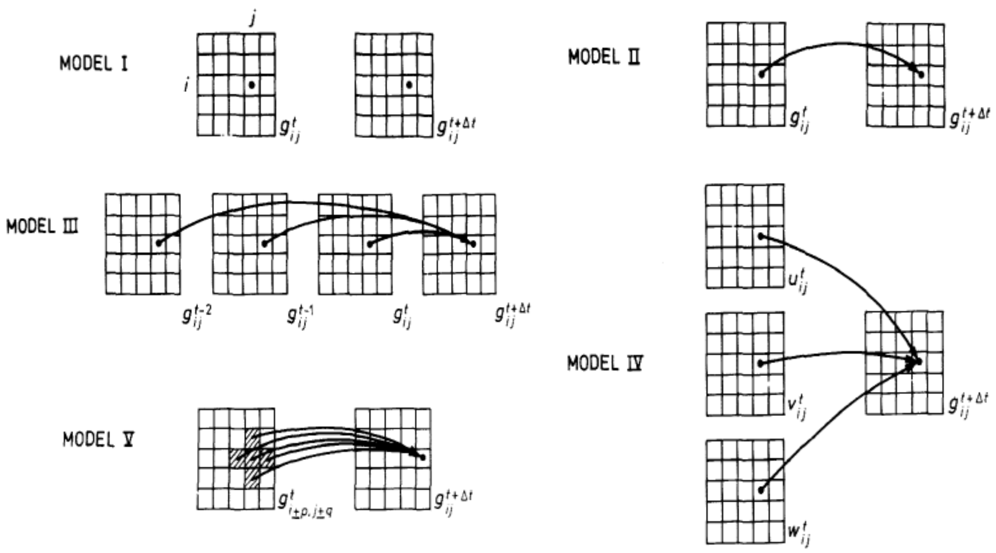


Figure 3.3 Classification of models of land use change (Tobler, 1979)

Transition rules F are defined closely to conventional CA definition. Tobler is more interested in studying geographic features of transition rules such as spatial isotropy and spatial stationarity. By spatial isotropy it is meant that the positioning of neighbours does not have any influence in the transition rule. By spatial stationarity it is meant that the same environment, the same neighbourhood, results in the same consequences; that is, the rules do not depend on where the location of the cell. These conclusions are, of course, very generic and considerably distant from real world geographic behaviours ground-breakings ground-breaking work was followed by a series of other studies that explored CA and its application to geography (Couclelis, 1985a, White and Engelen, 1993).

In the mid-1980s, Helen Couclelis continued the work started by Tobler in the research of CA for urban modelling purposes (Couclelis, 1985a). Her work has two main lines: the first one relates to the conceptual and theoretical linkage with the theory of complex systems; the second one relates to the exploration of possible uses in urban planning (White and Engelen, 1993). She establishes a parallelism between the set of transition rules of the Game of Life and urban changing phenomena. A 'live' cell could be interpreted as an urban zone that exceeds some kind of threshold in terms of a given urban function. Transition rule 1 (survival) would then guarantee that this cell would maintain this urban function if two or three neighbouring cells also exceeded that threshold. Transition rule 2 (death) would make this cell loose that urban function if four or more neighbouring cells also exceeded that threshold because the cell would suffocate or, on the other hand, if there was only one neighbouring cell exceeding the threshold (then the cell would die from loneliness). Transition rule 3 (birth) would settle this urban function in a cell if there were exactly three neighbouring cells that exceeded the threshold. A cell would remain dead if none of the precedent conditions were verified (rule 4, dead). However, the inherent simplicity of this formulation makes it inappropriate to simulate real world spatial phenomena. Couclelis identified a series of shortcomings on the ability of cellular models to simulate urban phenomena. Issues regarding space dimension and boundary problems are pointed out as the first limitations for modelling cellular worlds. The imposition of universal transition rules to an infinite regular cell space collides with an empirical interpretation of the method. A second limitation is related with the regularity of neighbourhoods. In order to be representative of real world phenomena, neighbourhoods should be different in shape and size for different cells. Cell regularity and spatial homogeneity within each cell are also issues that make classic CA formulation far from being geographically representative. Finally, it is also important to notice the assumptions of space and time invariance for transition rules as well as the system closure to external perturbations (other than stochastic behaviours) (Couclelis, 1985a). The solution relies on the consideration of relaxations that would enhance the ability of a cell-space approach to simulate real world phenomena without losing both its identity and its simplicity. Couclelis reformulated the structure of the Game of Life to obtain a simple geographic CA formulation, a generic structure that could be applied to a series of spatial problems (see Couclelis (1985a)). She also stated that this structure can be easily relaxed in order to improve simulation. Couclelis' formulation is independent from cell shape and size, and even from the existence of a cell space. The cell space concept can be generalized to the point where it describes any discrete time/space model representing components and their interactions (Zeigler, 1976 cited by Couclelis,1985). It is likely that any model of interest to geographers, whether aggregate or disaggregate, continuous or discrete, deterministic or stochastic, quantitative or categorical, can be expressed in the same language as the cell-space concept (Couclelis, 1985a).

Another important concept was introduced by Roger White and Guy Engelen: the concept of constrained CA (White and Engelen, 1993). Conventional CA are modelled with the explicit intention of being as general as possible. Because of this they have two main characteristics: (1) they are defined on a homogeneous cell space, overriding the variability of cell characteristics as they exclusively relate cells with their state values regardless of their location on the cell grid; (2) they are unconstrained, so that the number of cells in each state is determined endogenously by the application of transition rules to the current configuration of cells (White and Engelen, 1997). In their initial work, White and Engelen (1993) made some important assumptions that can be considered an approximation of CA to more likely urban behaviours. The first assumption is related to the consideration of a small number of hierarchically ordered cell states. There are four cell states, vacant (the lowest state), housing, industrial and commercial (the highest one). A cell in the vacant state can change to any other state but the inverse is not possible (thus the city can only grow, which is unlikely to happen). A cell in the industrial state can only change to commercial state. Another assumption is related to the use of neighbourhoods larger than the traditional Moore or von Neumann ones. The interaction between cells is then dependent of a larger area of influence, breaking up with the formal concept of local neighbourhood and introducing what is commonly named as action-at-a-distance. On a classic CA model changes of state must be locally influenced, discarding any action-at-a-distance (Batty et al., 1997). In this case, these relationships (depicted in Figure 3.4) are assumed non-monotonic and may present positive values (meaning an attractive relationship) or negative ones (a repulsive relationship). Finally, White and Engelen considered transition rules that are not based on a simple probability of change but on a composite measure of a transition potential. This transition potential is defined as a weighted sum that simulates the behavioural propensities of the actors who determine land use, as expressed in Equation 5:

$$P_{ij} = S\left(1 + \sum_{h,k,d} m_{kd} I_{hd}\right) \tag{3.5}$$

where P_{ij} is the transition potential from state i to state j, m_{kd} is the calibration parameter applied to a cell in state k at distance d, h is the index of cells within a given distance, I_{hd} is equal to 1 if the state of cell h is k and 0 otherwise, and S is a stochastic perturbation. The stochastic perturbation is calculated through the following expression:

$$S = 1 + [-\log(R)]^{\alpha} \tag{3.6}$$

where R is a uniform random distribution in the interval]0,1[and α is a control parameter for the adjustment of the size of the perturbation. This term has a highly skewed distribution so that most values are near one and much larger values occur only infrequently. The weighted sum in the transition potential is multiplied by S in order to simulate the stochastic behaviour of agents in each component of the transition potential.

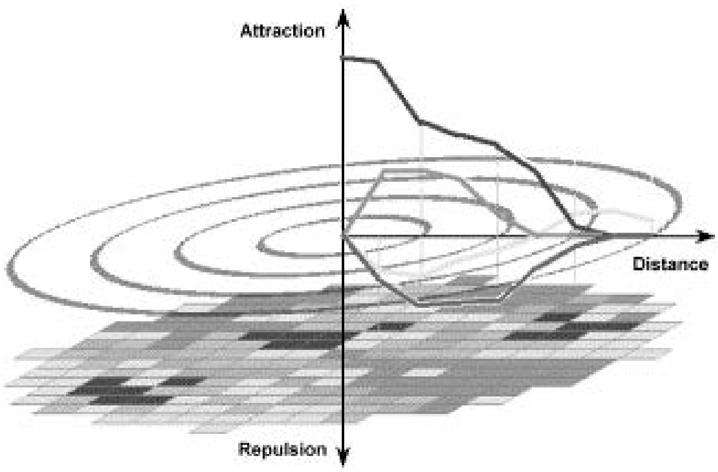


Figure 3.4 Neighbourhood interactions (Benenson and Torrens, 2004)

Another important evolution from the early CA models is related with the introduction of land use demand as an exogenous factor. In classic CA, the model evolves over time from a small set of cells to increasingly more complex systems through transition rules. The number of changeable cells is determined only by the change process itself. White and Engelen (White and Engelen, 1993) introduced an exogenous parameter that aims to simulate land use demand: there is a predetermined number of cells for each state that is considered in the transition procedure. If the total number of cells is not attained for a given state, there is a probability that converts cells in lower cell states to the given cell state until the demand is fully satisfied. The transition potential is calculated for every cell at every time step of the simulation and is ordered in a descending manner. The set of cells with higher potential is chosen for transition considering land use demand. This approach has been developed and improved, and was used in recent studies (White and Engelen, 2000, Barredo et al., 2003). Land use usually depends on three main types of factors: the inherent qualities of land itself, the effects of neighbouring land uses, and the aggregate level of demand for each land use (White and Engelen, 1997).

Another innovation introduced by White and Engelen (White and Engelen, 1993) is the consideration of fractal measures to assess model performance and simulation results. Several fractal measures are calculated for the set of theoretical cities generated by the model and then they are compared with known fractal measures from real world cities such as Cincinnati, Atlanta, and Houston. The fractal dimensions obtained for the theoretical cities were in line with the values observed for the set of real world cities, which indicates CA's ability to produce realistic simulations of urban growth.

It is also important to refer the work of Michael Batty on CA and urban studies. Batty is the author of a series of studies where CA is studied both from a theoretical and an operational point of view. His recently published book Cities and Complexity provides a framework for understanding cities as complex entities and presents CA (as well as agent-based simulation and fractal theories) as powerful tools for simulating complexity (Batty, 2005b). In his long work on urban simulation, Batty studied CA as a tool for exploring city complexity, mostly by formulating and implementing a series of CA models aimed to explore theoretical issues regarding the technique. He developed, with Yichun Xie, models for what they called "areal automata" to simulate urban structure, and for "linear automata" to simulate the growth of road networks with hierarchical structure (Batty and Xie, 1997). They then generalized these two types of models and developed an urban growth CA model based on cells and networks. These models were based on simple formulations close to the probabilistic formulation of conventional CA. The models also considered few land uses which shows their exploratory character. Later, a more sophisticated simulator was developed based on the work of Xie. The dynamic urban evolutionary modelling (DUEM) was aimed to capture urban complexity and urban dynamics. It was applied both to theoretical urban structures and to real urban areas, such as Amherst, New York (Xie, 1996). Batty also pioneered the coupling of CA and GIS as a means of developing visualization, which is increasingly important as a simulation tool (Batty et al., 1999, Batty, 2005b).

A series of more recent applications of urban CA are also important to the history of CA and urban modelling. However, it is important to first discuss the main evolutions of CA that are essentially based on relaxations of its four main components. The vast majority of these applications (if not all) introduced important innovations on urban CA pursuing the goal of improving their ability for simulating complex urban phenomena.

3.3.1 Main relaxations and evolutions

It was already mentioned that the simplicity of CA is one of their great attractions for urban modelling. From a simple set of rules operating over a simple cell structure it is possible to achieve complex forms from simple structures. But the classic formulation of CA is limited in its own formal definition, imposing the necessity of relaxing some of its basic components. It can be said that the notion “cellular automata” was used in geography, from the beginning, in a very broad sense, and not as a rigid formal scheme (Benenson and Torrens, 2004). The formal framework of CA defined by their most famous researchers as Ulam, von Neumann, Conway or Wolfram, based on a very simple formulation, is far from being able to represent real cities; many, if not all, urban CA bear little similarity with classic CA. One can question whether urban CA still are evolutions of classic CA or whether they are just cellular-based models (Torrens and O' Sullivan, 2001).

There are four major adaptations of the classic CA concept: (1) most applications to urban systems relax the local neighbourhood interaction to incorporate action-at-a-distance (considering this long range interaction as a global consequence of local spatial diffusion, thus reinforcing the application of classic CA); (2) it is hard to identify a scale for urban systems where everything is reducible to one activity in one cell; (3) the need of CA to meet plausible values of change rates; and (4) the use of GIS and map algebra (Batty et al., 1997).

Helen Couclelis pointed out two main characteristics that CA-based models must have: interactivity and realism (Couclelis, 1997a). Interactivity is an essential property that CA-based models must comprise. Being CA models of complexity, in which a small change in the initial conditions or transition rules may produce major changes in the results, they must allow the evaluation of small changes in model conditions to make sensitivity analyses possible. CA models should also incorporate good visualization techniques, not only at the graphic level but also in the statistical characterization of the results. The integration with GIS was pointed out as the next big step, because of the natural affinity between CA and raster images, and because of the potential of graphical visualization provided by GIS. However, the more complex is the CA structure, the more difficult it becomes to produce visualization. Another important feature of CA-based models is realism. Couclelis argues that no model based on the classic assumptions of CA – homogeneity, uniformity, universality – can claim good performance when applied to real world problems (Couclelis, 1997a). There are two main dimensions for model realism: realism with respect to data and realism with respect to model structure. The linkage between CA and GIS can be, once again, very profitable in order to guarantee data realism. Several CA models are already supported by GIS (Takeyama and Couclelis, 1997, Batty et al., 1999, Li and Yeh, 2000). Structure realism is achieved through the adaptation of the classic CA structure to the specific needs of simulating complex urban behaviours. CA ability to generate complex patterns from simple cell configurations and through small sets of rules is one of its main attractions, as it was already stated. But, in order to improve the feasibility of simulated urban landscapes, it is necessary to adapt the formalism of CA to the complexity of urban and social phenomena, adjusting the rigidity of classic CA components to the perception of the real world. Couclelis (1997a) identifies a series of relaxations that can be implemented in order to enhance the ability of CA models to correctly acquire the essential behaviours of complex urban dynamics. Figure 3.5 depicts a series of relaxations that can be implemented for all the four main components of CA.

The majority of CA models referred to so far are based on regular square cells forming an orthogonal cell lattice. The main reason for the use of these simple spatial structures relates to land use data availability from remote sensing maps. In order to enhance spatial representativeness, it is possible to forget this formalism and to use irregular cells as the spatial unit for CA. Tobler refers that there are some analytical advantages in considering the irregular spatial division of political jurisdictions (Tobler, 1979). However, Tobler states that the basic difficulty, of topological nature, relies on the fact that these irregular cells do not all have the same number of adjacent cells, thus their neighbourhoods cannot be defined by any simple notational scheme. There are very few studies based on irregular cells. The model developed by Vandergue *et al.* (2000) is cited by Ménard and Marceau (2005) as the only CA model that uses census tracts as cells. This approach has a great potential for linking urban form and reliable data, and is on the basis of the CA model presented in the next section of this chapter. Semboloni (2000) used Voronoi polygons to model urban growth considering cell division. O’Sullivan (O' Sullivan, 2001b, 2001) developed another CA-based model coupling CA and graph theory. Benenson and Torrens

(2004) refers the existence of theoretical studies that used triangular and hexagonal cells (Gerling, 1990, Eloranta, 1997).

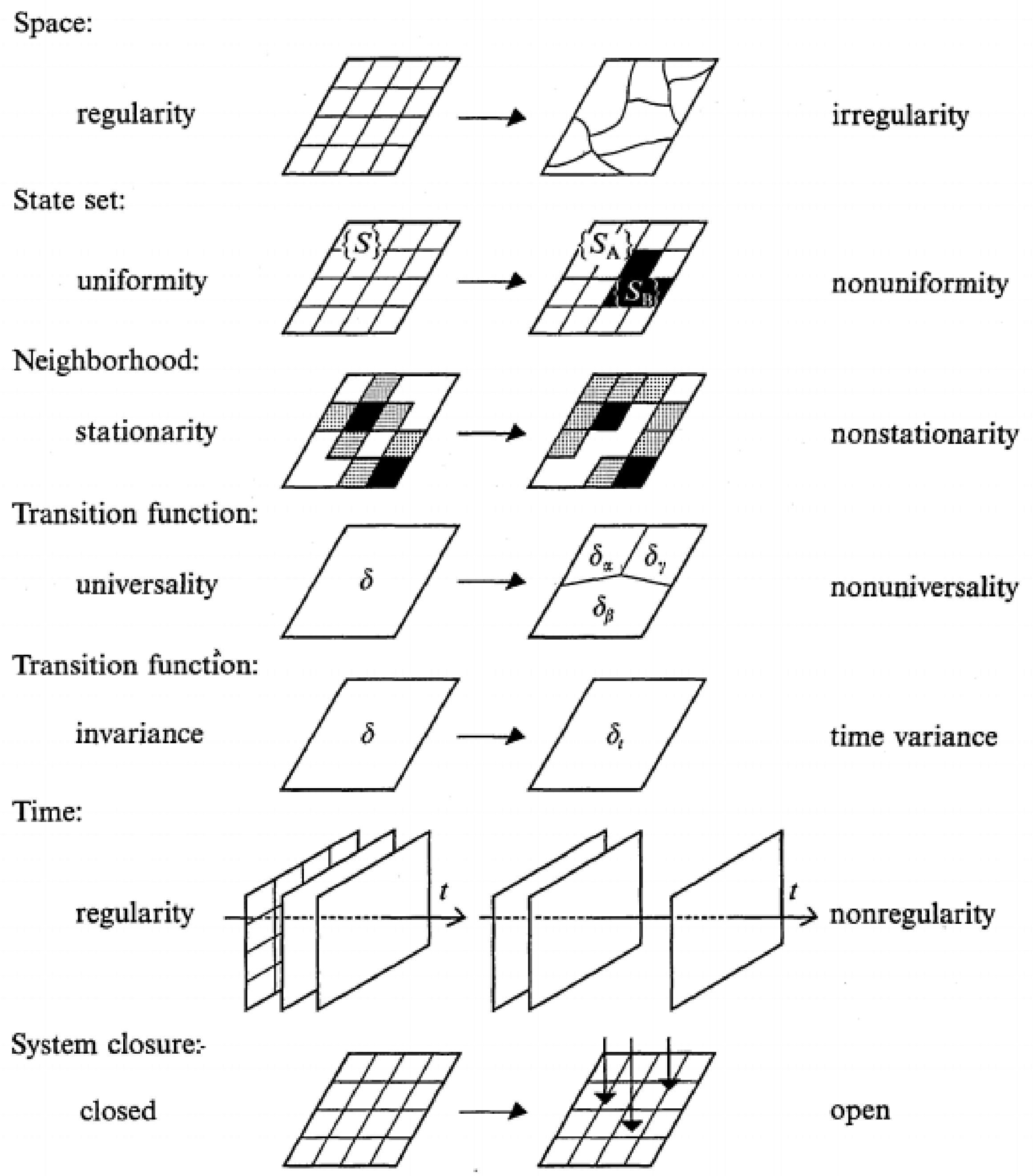


Figure 3.5 Possible CA relaxations (Couclelis, 1997b)

Cell state sets can also be adapted to simulate different land uses in different parts of the simulated area. This modification has some implications not only in the definition of neighbourhoods but also in the definition of the set of transition rules.

The concept of neighbourhood can also be modified. There is a problem of representativeness in the first place. The importance of the neighbourhood is that it defines the geographical domain of influence (Tobler, 1979). The fact that there are different concepts of neighbourhood must be taken into account: the sense of neighbourhood of residents of urban areas and residents of rural areas is different, therefore it may be useful to consider the shape, size and, form of a neighbourhood as a function of the location of its central cell, giving special attention to neighbourhoods in boundary areas (Tobler, 1979).

Neighbourhoods such as von Neumann's and Moore's are particularly suited to physical phenomena; human interactions are more likely to be explained by wider areas of influence (White and Engelen, 2000). Physical and topographic constraints can also shape the sense of neighbourhood, particularly in larger urban areas. Neighbourhood structure may also vary in time. There are some studies in the field of tissue and leave growth that considered the creation of new cells within a given neighbourhood (Lindenmayer, 1968). Semboloni (2000) also assumed that cells could be divided, thus creating new neighbourhoods. Ménard and Marceau (2005) produced an interesting study on spatial scale sensitiveness. They gathered information about the neighbourhoods used in several CA models from 1993 to 2003. They defined CA's spatial scale as a set of three components: spatial extent, cell size and neighbourhood configuration. They focused their study on the analysis of different cell sizes and neighbourhood configurations for a simple two state problem. Small variations in cell size can produce significant variations in results when a given scale threshold is exceeded. The authors state that CA models are not sensitive to variations of neighbourhood configuration. Spatial scale sensitivity affects CA models were the cell actually represents a portion of the geographic space. Ménard and Marceau (2005) also stated that, although scale issues are supposed to be considered as a design choice that will not drastically vary during the simulation process, modellers must pay more attention to the subject, because some of the features that influences neighbourhood are effectively dependent on scale choice.

Transition rules also underwent major developments from their initial formulation. The main component of CA is the transition rule (Torrens, 2000). The first and more significant evolution regarded the incorporation of stochastic perturbations in transition rules. Deterministic behaviours are not suited for the simulation of complex urban phenomena because they can be considered a result of complex interactions between agents. For this reason, many CA models have incorporated stochastic perturbations in their transition rules (White and Engelen, 1993, Clarke et al., 1997, Barredo et al., 2003). Transition rules can also be considered non-static, varying through time. Their universality, an assumption of the classic CA formulation, can also be relaxed, as different areas and land uses present different change behaviours. There may be also a distinction regarding the way transition rules operate over time. They can be applied sequentially, updating cell states one after another, or in parallel, updating all cells at the same time (Benenson and Torrens, 2004). Von Neumann's self-reproducing CA presented an asynchronous behaviour with the cells being updated considering the previous updates generated by the automata. Conway's and Wolfram's approaches were based on synchronous behaviours, with the entire cell set being updated at the same time. There are two main processes of asynchronous updating. The cell set can be updated at a moment in time after a predetermined order according to some cell characteristic. Another method is based on a probability of change in a given moment that is a function of the time the cell has to wait for changing.

Finally, formal CA can be considered as a black box that processes input data towards an output without any interference from the outside world, that is, they can be considered as a closed system. Urban systems are too far from being close; in fact, an urban system is one of the most opened systems that can be found. Therefore, the assumption that an urban system modelled by CA can experience exogenous interference – say stakeholders' actions or political decisions – enhances their ability to simulate urban complexity and, consequently, their representativeness. Many models deal with this relaxation by introducing stochastic perturbations that only occur when certain exogenously defined thresholds are overcome (Clarke et al., 1997).

Relaxations are needed to improve CA's ability to simulate complex urban phenomena. The use of real, disaggregate data made possible the relaxation of any one of all the assumptions of classic CA, allowing the models to better fit the complexity of cities (Couclelis, 1997a). However, excessive relaxation of traditional CA assumptions may increase exponentially the difficulties to understand the outcomes of a model, in a comeback to widely criticized large-scale simulations (Couclelis, 1997a).

3.3.2 Applications of CA in urban change problems

One of the most widespread CA models is SLEUTH, developed by Keith Clarke to reproduce and predict urban growth (Clarke et al., 1997, Candau, 2000, Clarke, 2002, Silva and Clarke, 2002). Its name is an acronym for Slope, Land use, Exclusion, Urban, Transportation and Hill Shade. SLEUTH has two main modules (Clarke, 2002): first, an urban growth model; second, an embedded land use model that uses information from the urban growth model. This model requires six GIS-based inputs for visualization, used in image format: urbanization, land use, transportation, areas excluded from urbanization, slopes, and hill shading. Urban areas are required for four different time periods for calibration purposes. Urbanization is the product of an urban seed file and at least two road maps that interact with a slope layer to allow the generation of new urban centres. It considers a regular grid space, a neighbourhood of eight cells and only two cell states: urban and non-urban. It operates five sequential transition rules: (1) diffusion, (2) breed,

(3) spread, (4) slope resistance, and (5) road gravity. The growth rate depends on five different factors. The diffusion factor determines the overall dispersion of the distribution of single grid cells and of the movement of new settlements outward through the road network. The breed factor determines how likely a newly generated settlement is to begin its own growth cycle. The spread factor controls how much outward "organic growth" expansion takes place within the system. The slope resistance factor influences the likelihood of settlement existence on steeper slopes. The road gravity is a factor of attraction of new settlements onto the existing road system if new areas fall within a given distance of a road (Clarke, 2002). The operation of these factors leads to four different types of urbanization. First, spontaneous growth: any non-urban cell can be urbanized according to a probability inversely proportional to cell slope. Second, generation of new diffusion centres: each spontaneously urbanized cell can become a new spreading centre if it has a given number of neighbouring urban cells and reaches a probabilistic threshold defined as a model parameter. Third, diffusion at the edges of urbanized areas: there is a fixed probability (another parameter of the model) that allows an edge cell to become urbanized given a certain number of neighbouring urban cells. Fourth, road-influenced diffusion: a new spreading centre is chosen given its distance to the road network. It can be dislocated along that road in a randomly selected direction for a given distance (another parameter of the model) for a new location where it can "paste" development. Two randomly chosen neighbouring cells would then change state. An important improvement of this model is the consideration of self-modification rules aimed to modify the model's behaviour over time giving it the ability of identifying different periods of intensive growth or of little or no growth.

After the first phase of urban growth modelling, SLEUTH will assign land uses to the new urban areas obtained from the first phase of the model. This assignment is made through an embedded model named Deltatron. The main assumption is to consider each cell as a Deltatron, an urban entity that is associated to only one cell state (or land use) (Benenson and Torrens, 2004). These entities evolve during simulation in a way similar to the one described for urban growth, through four stages. Firstly, cells are selected at random as candidate locations for land use change on the basis of how much urban growth has taken place. Each newly urbanized cell is assumed to induce a potential change in land use and, as a result, determines whether the selected cell will keep the same land use or change to another; this is made through the consideration of a probability that depends on historical change and cell slopes. If a transition occurs and the cell is not a Deltatron already, a new Deltatron is created. Cluster dynamics are defined as an aggregation process of these new Deltatrons and the associated land use transition. The newly transitioned cell acts now as the land use aggregation centre. This process behaves closely to the “organic” growth described above. Age is also considered to characterize Deltatrons: as time goes by, Deltatrons get older and are eliminated when their age reaches a given threshold. SLEUTH has the credit of introducing the concept of evolving transition rules with the aim of improving the model’s ability to retrieve past behaviours. It is also a straightforward model that can be easily calibrated to different urban and regional areas with a small set of calibration parameters. There are already an important number of applications not only in the United States but also in Europe, Africa and South America.

Another innovative approach was proposed by Li and Yeh (2001b), with the basic model structure is depicted in Figure 3.6. They formulated a CA-based model that uses artificial neural networks (ANN) to simulate urban change. It has been applied to the Pearl River Delta, China.

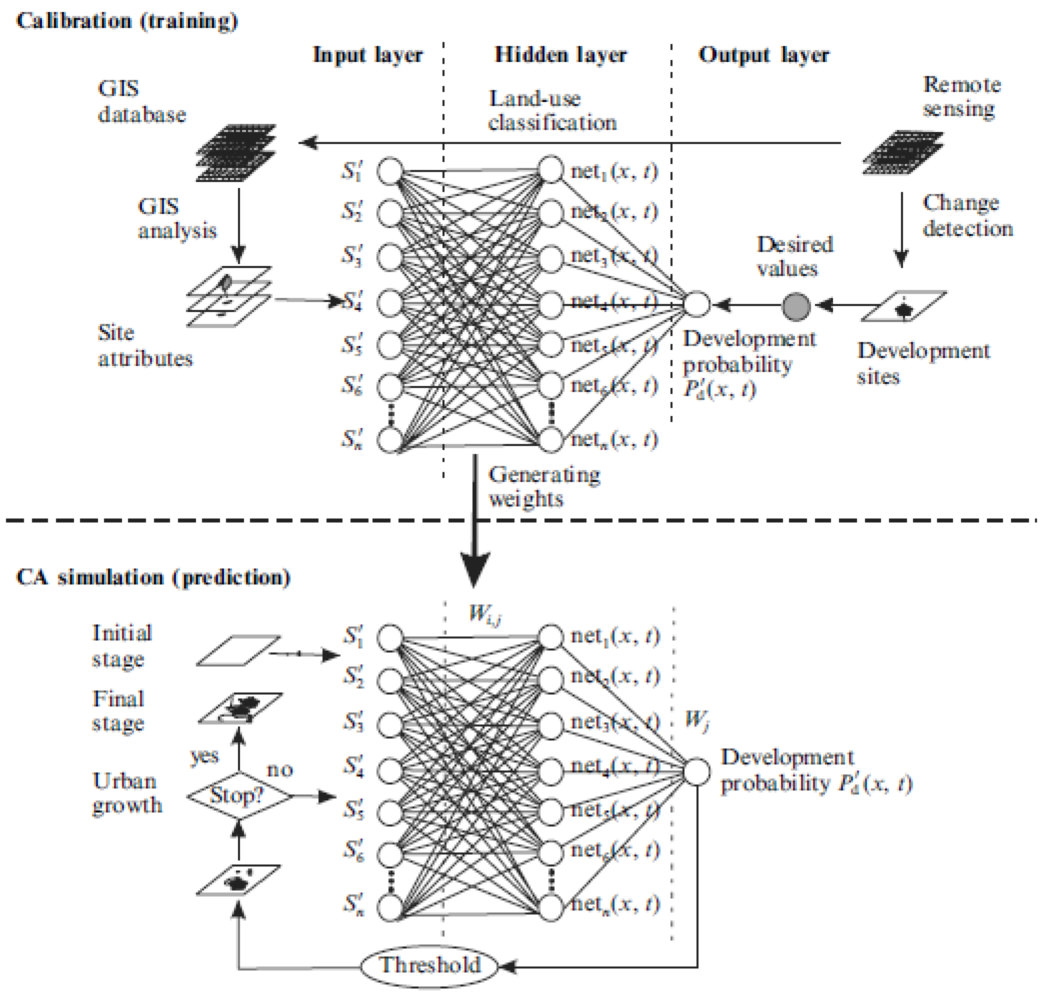


Figure 3.6 CA model based on artificial neural networks (Li and Yeh, 2001b)

The set of parameters is obtained automatically by the ANN method, which is made more robust because it uses a back-propagation training procedure. The authors consider that the traditional CA formulation has serious problems in obtaining consistent parameters values. The method is believed to be suited for dealing with complex interactions between dependent variables without user involvement, a major issue in urban modelling. The model is capable of analysing relevant data, and to eliminate noise and other redundant data. Another important development of this model is related to transition rules: they are obtained from the ANN training process and not from user definition. The model only needs to be fed with training data. It is also capable of dealing with uncertainty by simulating alternative developments through network training, integrating potential interventions exogenous to the change process which are very difficult to be deducted from historical data.

David O'Sullivan (2001b) proposed a model that couples CA with graph theory to enable research into relationships between spatial structure (represented by graphs) and urban dynamics (simulated by CA), as depicted in Figure 3.7. The model is based on a partition that needs not be space-filling, and overlapping spatial elements might be used. In order to study urban phenomena, the entities forming the basis of urban morphology (buildings, blocks, streets, census blocks, and administrative/planning zones) are a natural set of spatial elements to be used as graph vertices.

Edges in the graph represent some sort of relationship between vertices, so that any relationships relevant to the model being developed might be used. The neighbourhood of each vertex consists of a set of adjacent vertices. Cell states may be defined in a suitable way, considering land use classes as large as necessary to correctly simulate reality. The graph/CA model was applied to study gentrification in small local neighbourhoods in central London. Data availability and quality were considered problems for using this approach to micro-scale modelling. The dynamic behaviour of complex models was pointed out as an important shortcoming to the application of micro-scale modelling to gentrification problems.

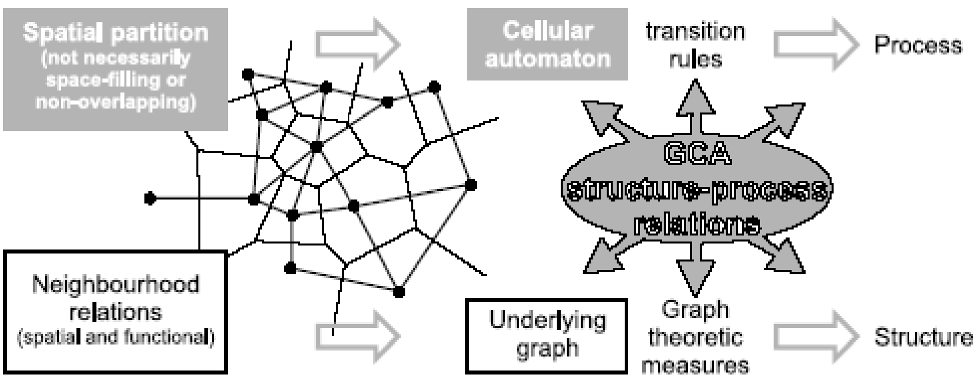


Figure 3.7 Graph-CA model concept (Source: O'Sullivan, 2001b)

Barredo *et al.* (2003) developed a constrained CA model named MOLAND, which is an integrated platform that includes a series of macro-scale models for simulating the natural, social, and economic sub-systems that are the main drivers of land use evolution at this scale. These models are coupled with a classic CA model for land use allocation at a local scale, following the approach of White and Engelen (White and Engelen, 1993). MOLAND uses the concept of transition potential introduced by these authors to build a model based on a regular cell grid with a radial neighbourhood of 172 cells. They also considered a neighbourhood effect materialized through a user-defined weighting factor matrix, establishing quantitative values for attraction and repulsion, similar to the concept depicted in Figure 3.4. The model was applied to a series of regions in Europe and also Africa, particularly for supporting risk assessment analyses. One of the case studies was Dublin, for which a thirty-year long historic period (1968 to 1998) was used for calibrating the simulation. Using fractal measures and contingency matrices as goodness of fit functions, the model was able to achieve interesting results. Simulated fractal measures for each land use were quite similar to real values, and the values for the statistical indicators derived from the contingency matrices can be considered satisfactory.

3.4 Future trends in cellular automata

Cellular automata are an important subject of interest in urban simulation. The concept of CA became even more important with the great developments that multi-agent systems (MAS) experienced in the past years, by considering cells as spatially located agents with well-defined behaviours. Cells are non-moving agents that are properly suited for simulating a series of phenomena that are located in space. The consideration of cells as agents requires more research on some of the main concepts of CA, such as scale, cell form, neighbourhood, and transition rules. Some hints are pointed out below for the pursuit of these new research paths.

3.4.1 Multi-scale cellular automata

Different urban phenomena can be observed and modelled considering, on the one hand, cities as parts of larger regional systems and, on the other hand, cities themselves, where the driving forces of urban change depend on local variables. The issue of scale is currently under debate and some models already face the problem of inter-scale interaction (Kocabas and Dragicevic, 2006, Benenson, 2007, White, 2007). The use of both regional and local scales is considered useful to correctly simulate a wide set of complex urban phenomena that occur on different spatial scales. Considering that these demographic and economic relationships depend largely on neighbourhood conditions, and that large scale cells can be observed with a useful amount of reliable data (counties, municipalities, cities), it is clear that a CA model may be used to simulate urban evolution at a regional, therefore macro-scale. At this scale, the assessment of aggregate land use demand – for housing, industrial, or tertiary land uses – through population and employment growth can be more representative of urban growth than the disaggregate amount determined for each land use, which is traditionally used by common CA models. In opposition, at a city/neighbourhood scale – the local scale – land use growth outcomes from the distribution of each land use (considered in a disaggregate level). In fact, land use demand can be estimated as the amount of land necessary for each land use considering population and employment growth. Then, a set of disaggregate land uses can be assigned to different cells in order to meet land use demand. This multi-scale approach is the consideration of two levels of simulation (as depicted in Figure 3.8): the regional level, for which the aim is to assess land use demand from population and employment evolution through time; and the local level, for which the aim is to assign different land uses to different locations (cells or parts of cells) in order to meet land uses demand.

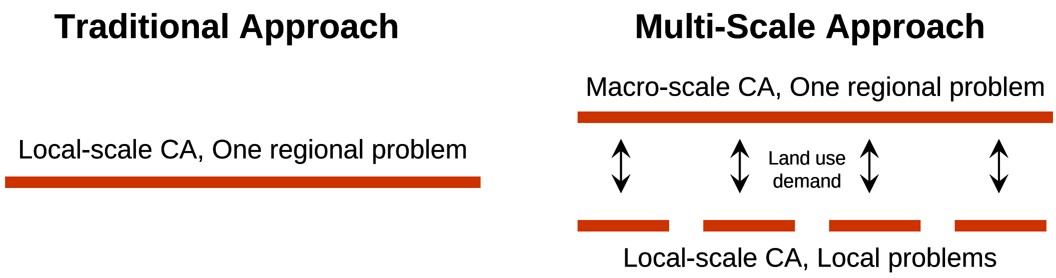


Figure 3.8 Multi-scale CA structure versus traditional CA approach

3.4.2 The use of irregular cells

Common CA models operate with identical square cells that are usually obtained from remote sense maps (White and Engelen, 1993, Clarke et al., 1997, Barredo et al., 2003, Li and Liu, 2006). This approach was the natural evolution from the 2D mathematical CA models and is useful for modellers because it allows the collection of reliable historical data on land use, irrespective of the ability of this framework for fitting real-world processes (Benenson, 2007). However, these cells are not directly related to urban form because of their regularity. The use of irregular cells has implicit the idea of capturing the natural irregularity that can be observed on the structure of a large majority of urban areas around the world. The number of CA models that use irregular cells is small (Semboloni, 1997, Vandergue et al., 2000, O' Sullivan, 2001b, Stevens et al., 2007). The potential of this approach derives from the possibility of combining urban form and reliable data. At a regional scale, municipalities or other intermediate administrative units will in principle be appropriate cells; at a local scale cells should be as close to urban structure as possible. Census blocks can be a good option (they proved to be feasible for the CA application to small urban areas presented before) since they are normally drawn considering the urban structure, and contain extensive and reliable information. There is indeed a large amount of processed data of various types both at a regional and a local scale. The possibility of crossing reliable data and spatial structure is the main gain obtained from using irregular cells.

3.4.3 Dynamic neighbourhoods

Neighbourhood is a critical issue for every type of spatial models, and in particular for CA. Neighbourhood is commonly (if not exclusively) considered by the strict concept inherited from the mathematical formulation of CA. This concept is based only on the consideration of a set of physical neighbours to one cell: these neighbouring cells can be those which are directly connected to the cell considered, or they can be the group of cells that are within a given range from that cell (White and Engelen, 1993, Benenson and Torrens, 2004, Ménard and Marceau, 2005). But this mathematical perspective is far from being representative of how cities work. The concept of neighbourhood must be able to reproduce how agents interact, considering both spatial and functional levels of interaction. At a regional level, neighbourhood is influenced by both these levels of interaction. A municipality is strongly influenced not only by its directly connected neighbours – in terms of employment and trade – but also by the main functional centres of the region, where administrative and economic decision centres are located. At a local level, this dual influence is also observed. The choice of location for a given land use is influenced both by the surroundings (a good example is the relationship between residential and industrial land uses) and by the distance to the main functional and employment centres in the city. Neighbourhood must shift from the concept of a limited area which remains constant in time to a larger and possibly disconnected part of the territory, varying in time. A cell can have a discontinuous neighbourhood that includes its immediate neighbours at a certain extent and some distant functional centres (regional capital, main regional cities) as depicted in Figure 3.9. This neighbourhood can vary during the years, enlarging or decreasing proportionally to the potential of attraction of the cell. This concept of neighbourhood implies the consideration of spatial and functional interaction between cells.

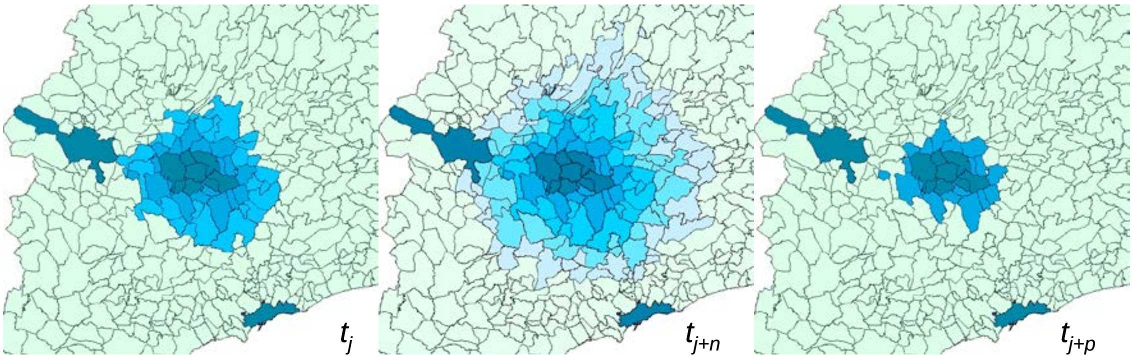


Figure 3.9 Dynamic neighbourhoods

The main goal is to enhance neighbourhood representativeness at both regional and local scales. There are physical and ecological barriers that limit or break the shape of a neighbourhood. At a regional scale, a mountain, a river, or a delta can act as a physical barrier for many types of interactions. The same occur at a local level with topographic features, heavy urban infrastructures (road and railways), and city form acting as barriers for defining the perception of neighbourhood.

3.4.4 Adjusting transition rules to different scales of interactions

Transition rules are also an important component of CA. They define the way a CA model evolves through time and they are intimately related to the way the model is able to correctly acquire behaviours that are the drivers of the simulation. Classic CA uses probabilistic sets of transition rules (Wolfram, 1984). Although this probabilistic approach has been used in some theoretical urban CA models, transition rules are commonly designed to reproduce complex urban behaviour (White and Engelen, 1993, Clarke et al., 1997, Barredo et al., 2003). These rules can also be used to reproduce planning regulations that constrain land use change and demand. A multi-scale approach requires some attention on the definition of transition rules for both scales. At a regional level of analysis, it is important to notice that the goal is to simulate the macro interactions that are observed within a region. These interactions can be assessed using macro-scale indicators of population, employment, and commuting flows, for example. Cell states should relate to these indicators and can be defined as a degree of urbanization taken in aggregate or disaggregate form. A binary state of urbanization, occupied or non-occupied, can be used; an alternative approach can use a disaggregated degree of urbanization, in percentage of the available land. It is also possible to consider the main land uses (residential, industrial) in an aggregate or disaggregate way. At a local scale, the goal is to simulate the distribution and growth of land use using a more disaggregate classification. These phenomena depend on a series of factors related with land suitabilities, accessibility, land use demand, and land management policies, among many others. At this scale, transition rules will relate to measures of urban potential for each land use, considering all the interactions between all the land uses, observing planning constraints that can be derived from planning regulations. Cell states are the traditional set of disaggregate cell states that are commonly used in classic CA models.

3.4.5 Considering policy testing in CA

Policy testing is widely recognized as an important feature that must be taken into account in the development of urban simulation. The use of simulation by itself is often considered redundant because it tends to produce results that are sometimes considered futile by planners and decision makers. Simulation must be able to incorporate the practical needs of planners and of the planning process. There are different degrees of implementation of simulation-based methodologies across the world, which relates with different contexts of planning, regulation, and available information. Simulation is oriented for understanding and reproducing real phenomena in a controlled environment, aiming to explain how these phenomena work and how can they be dealt with. However, these features do not provide planning with the necessary tools for dealing with issues that are strongly influenced by uncertainty. It is necessary to enhance the ability of models for using the parameters that characterize complex behaviours, allowing the consideration of policy testing on the generation of simulated scenarios and evolutions. The consideration of policy testing in CA is made using transition rules to reproduce policy constraints. This process is based on the identification of candidate policies to be tested by the model, by considering transition rules that can incorporate the parameters that control those policies. Its implementation must be an ongoing process that is expected to last even after the end of the model development stage: every time a new policy is set to be tested it should be possible to change the set of transition rules in order to meet specific simulation needs. This type of simulation is strongly dependent on the historical data gathered for a given problem. Every simulation technique can be considered captive of this data, reducing its ability of forecasting an uncertain future. This is probably the most important shortcoming that is pointed out to modelling by common planners. Therefore, it is important to develop modelling tools that are able to correctly identify historical trends and to break them under given conditions. Major urban transformations are usually the result of single decisions localized in time as the large urban renovation operations or the organization of important sporting events (for example the Olympic Games). These types of transformations are almost impossible to capture by any model because they cannot be derived from historical data. CA work with a set of transition rules that are calibrated considering the historic evolution of urbanization. The model must be able to generate new transition rules after observing a trend that cannot be explained by past evolution. Another possibility is the calibration of transformation thresholds. An illustrative example is the change of a cell to land uses that are forbidden by planning regulations. Once those thresholds are crossed, a given cell should be allowed to change to cell states that were forbidden before.

3.4.6 Land suitability indicators for cell differentiation

Land suitability is also an important issue to attend. Different land uses demand different land suitabilities for choosing a given location. Although this is not a classic CA component, it assumes great importance when one is dealing with constrained models. A model must be able to correctly differentiate two neighbouring cells in terms of their ability for allocate different land uses, which is a crucial issue for improving representativeness. At a regional level, the analysis is made from an aggregate point of view and land suitability is seen as a measure of comparison for general environmental quality or environmental protection policies. It is important to capture the influence of general environmental and physical characteristics in location choice, both for residential and for non-residential land uses. There are municipalities that have more demand for residential land uses due not only to their environmental characteristics but also to the distance to large industrial districts or polluted areas. At a local level, the comparison between land suitabilities for every land use is decisive for assessing demand for different and/or competitive land uses. Land suitability is an important constraint to the occupation of a given land plot by a given land use. Therefore, at this level of analysis it is imperative to develop a robust set of land suitability indicators that include physical features and, possibly, measures of levels of service for basic urban infrastructures as water supply or solid waste collection.

3.4.7 Accessibility measures for CA models

Accessibility is strongly linked to land use. Any attempt to simulate complex behaviours that occur within an urban system must take into account accessibility, as it may be considered as one of the most important drivers of urban growth. Accessibility is also strongly dependent on the scale of analysis. In order to correctly simulate accessibility conditions and evolution throughout time, different transportation modes must be considered and their scopes of influence must be attended. Air and maritime transportation modes have a regional scale of influence, as the number of network nodes located in a given region is always only a few. On the contrary, road and rail transportation modes have both a regional scope (regional highways and roads and railroads systems) and a local one (city road network, subway rail systems). Future research will probably focus on the development and testing of different measures of multi-modal accessibility that can be used as an input for a CA model at different scales of simulation.

3.5 Conclusions

Cellular automata (CA) models are still in the front line of urban simulation techniques. After more than two decades of intensive research on geographic CA, following the pioneer work of Helen Couclelis (Couclelis, 1985a), CA have already a solid scientific background that projects its use into the future of simulation. Their intrinsic spatiality guarantees that the concept has an important role to play not only as a modelling tool by itself, but also coupled with other simulation concepts and techniques, such as multi-agents modelling (MAS). Furthermore, this spatiality comes useful for developing decision support tools for assisting common planning processes, because CA are able to easily operate over land use maps, which are one of the most wanted tools for planners and decision-makers. This feature fosters the use of CA-based simulation within the GIS framework. But to guarantee that CA continues to appear as an interesting concept for both planners and modellers to develop modelling-based planning approaches, it is crucial that the concept evolves towards a more realistic level of simulation by opposition to the more strict mathematical formulation traditionally followed. The research on CA's components is fundamental to achieve this goal. Cells, neighbourhood and transition rules still have an important margin for research and development. There are several issues that depend on innovative computational approaches. Calibration, a key issue in modelling, can also be improved through the use of more sophisticated algorithms. New challenges are posed to CA and open new perspectives on their use in simulation. The coupled use of CA with MAS is probably the most obvious evolution of CA, creating the new concept of Geographical Automata Systems (GAS) proposed by Benenson and Torrens (2004). This integration provides a very powerful tool for modelling the physical and socio-economic phenomena that are the main drivers of urbanization, by linking spatial and non-spatial agents, enhancing our capacity for capturing and understanding complex behaviours. This creates new possibilities in forecasting future evolutions of urban growth by ensuring higher degrees of realism and detail. CA's simplicity and spatiality ensure them, as a standalone concept or coupled with other modelling techniques, a promising future.

4 A First Cellular Automata Model: Model Development and Calibration

Models in general are abstract formulations of a reality they are designed to describe. Models have different extents of adhesion to that reality, from a more theoretical and conceptual perspective to a fully-fledged degree of operationally. The extent to which models respond to this degree of abstraction must be accessed in terms of its applicability and in terms of the techniques and the outputs of their calibration.

This chapter addresses the issues of applicability and calibration of cellular automata (CA) models and has two main goals. The first goal is to assess the applicability of a CA model to the study of urban change in small urban areas. CA are usually applied to the study of regional or metropolitan areas (White and Engelen, 1993, Clarke et al., 1997, Silva and

Clarke, 2002), discarding the consideration of local problems outside their geographic contexts. The second goal of this study regards the use of irregular cells on a CA approach. There is only a small group of studies that developed CA models considering irregular cellular fabrics (Semboloni, 2000, Vandergue et al., 2000, O'Sullivan, 2001). This particular characteristic is of great importance when traditional regular cells, obtained from satellite images, do not represent well the spatial structure of the territory. Irregular cells based on the urban structure can be considered as “natural spatial cells”. Census tracts are drawn considering the urban structure, holding at the same time structured and reliable information on demographics and construction. Therefore, they are a good spatial partition when irregular cells are considered.

The chapter is organized as follows. In section 4.1 a brief literature survey on CA and focusing on the calibration of CA models presented. Section 4.2 is dedicated to presenting the CA model developed in the research. Section 4.3 discusses the definition of performance measures, followed by the description of the calibration procedure in section 4.4. Section 4.5 presents the set of theoretical test instances used to assess model performance and in section 0 these results are presented and discussed. Finally, in section 4.7 some conclusions are drawn.

4.1 Literature overview

The concept of cellular automata was introduced in the late 1940s by John von Neumann and Stanislaw Ulam, who were facing (independently) the problem of devising sets of mathematical rules for simulating the self-reproduction and evolution of biological systems. The spatial nature of CA led Waldo Tobler to pioneer their introduction in urban studies (Tobler, 1979). This breakthrough was contemporary of a very turbulent period in spatial modelling, after Lee's criticisms on the development of large scale models (Lee, 1973). Because of this significant, although brief "dark age", CA was kept aside from the scientific spot light until the mid-1980s, gaining a new breath with the advent of faster and more capable processing capabilities along with the development of more powerful database management tools. Tobler and also Helen Couclelis were responsible for giving CA the necessary scientific background for their application in geography and urban studies (Couclelis, 1985a). They were followed by many other researchers in different parts of the world and many theoretical and operational models were developed. Couclelis (1997a, 2005), Batty and Xie (1994), and Batty (2005b) worked on theoretical issues regarding CA application to urban studies. White and Engelen (1993, 1997), Batty and Xie (1997), and Clarke et al. (1997) worked on the application of important evolutions of CA to real world problems. CA also benefited from the contemporary development of the and GIS concepts and their coupled use was approached by, for example, Wu (1998) and Batty et al. (1999). Since then, a series of papers were published using CA models as the core method to address urban growth (Semboloni, 1997, Silva and Clarke, 2002, Barredo et al., 2003, Ward et al., 2003) The use of irregular cells is very small considering all the research made on CA. Semboloni (2000) used Voronoi polygons to model urban growth; O'Sullivan (O' Sullivan, 2001a, b) combined CA and graph theory generating a set of neighbourhood scale irregular cells; recently, a very interesting work on high resolution irregular CA was published, in which the cell is a single land parcel (Stevens et al., 2007).

Calibration always was an important research issue in the development of CA models. As important as the conceptual development of a model, calibration is aimed to ensure the necessary connection between simulation and reality. CA model calibration has been a subject of different approaches using different types of procedures, from sensitivity analysis to optimization-based methods. SLEUTH (Clarke et al., 1996, Candau, 2000, Silva and Clarke, 2002, Onsted and Chowdhury, 2014, Sakieh et al., 2015) is calibrated through a two-step procedure: first, a visual calibration oriented for a broad parameter definition and debugging; second, a brute force calibration procedure where multiple runs are produced in order to generate enough model data to statistically compare reference data. Li and Yeh (2002) and Li et al. (2013), Almeida et al. (2008), Basse et al. (2014) and Altartouri et al. (2015) coupled CA models with artificial neural network formulations to calibrate transition rules and other components of the CA models. White et al. (1997), Barredo et al. (2003) and Li et al. (2014) used basic sensitivity analysis to calibrate the weighting parameters for the spatial interactions between land uses.

4.2 Model presentation

The CA model presented in this chapter has a simple structure that derives from the classical formulation of CA with the consideration of constrained land use demand, following the concept introduced by White and Engelen (1993). The model operates over an irregular cellular fabric obtained from census blocks. Cell states are classified into a finite set of aggregated classes of land use. Land use interactions take place within a variable neighbourhood that is determined through model calibration. Transition rules incorporate planning regulations and simulate land use change based on a composite transition potential that takes into account cell accessibility, land use suitability, and neighbourhood interactions within the cell neighbourhood. The time step can be defined by the user. Land use demand is determined through the evolution of population and employment densities over time. The model was implemented from scratch using the Visual Basic 6 programming language to run on Microsoft Windows.

4.2.1 Cell structure

The model uses irregular cells to model urban form. The main goal is to explicitly introduce in the model the capacity of dealing with existing data that is used in policy making and spatial planning and use this data to process the spatial dynamics of urban change. Irregular cells are based on census blocks or equivalent units used to collect important datasets as the demographic censuses or employment data. These irregular cells are design by statistical offices to cope with the effort of collecting data, thus including the usual constraints of the natural and the built environment in its form. This way, the model effectively combines urban form with data that is linked to that form.

4.2.2 Cell states

The model is designed to work with different finite sets of cell states. Cell states represent an attribute that illustrates the way evolution takes place in the system. Common applications of CA models focus on land use as the main cell state and this model follows that logic. In this application the model makes use of a set of six aggregate states (or land use classes), which are intended to represent the whole set of urban land uses: urban low density (UL), urban high density (UH); industry (I), urban expansion (XU), industrial expansion (XI), and highly restricted uses (R) .States UL and UH are comprehensive classes of urban land uses, differing only on building density. They include residential areas, shopping and office areas, mixed-uses areas, streets, and open spaces. States XU and XI represent non-occupied areas where urban and industrial land uses can be located, respectively. Cells classified in any state but R are identified as active cells: these cells can change state under a given set of conditions. Cells classified as R, which correspond to land classified as agricultural reserve, ecological preservation area, etc., are identified as inactive cells: they can influence land use dynamics but their state cannot change.

4.2.3 Neighbourhood

The model uses a circular neighbourhood with a radius δ that is a calibration parameter. Although the use of circular neighbourhoods is quite common to several CA models (White and Engelen, 1993, Barredo et al., 2003), neighbourhood was always defined as an exogenous input which is purely user defined. Its value influences, and is influenced by, all the other parameters at stake.

4.2.4 Transition rules

Transition rules are based on a measure of state transition that is calculated through a potential that reflects the propensity/capacity of each cell to change state in each time step. This potential is a weighted value of land use suitability, accessibility, and neighbourhood effects given by the following expression:

$$P_{i,s} = \left(\nu_P \times S_{i,s} + \chi_P \times A_i + \theta_P \times N_{i,s}\right) \times \xi, \forall i \in \mathbf{C}, s \in \mathbf{S} \tag{4.1}$$

where, for each cell i from the set of cells **C**, and for each state s from the set of states **S**, $P_{i,s}$ is the transition potential for state s of cell i, $S_{i,s}$ is the land use suitability value for state s of cell i, A_i is the accessibility value of cell i, $N_{i,s}$ is the neighbourhood effect for state s of cell i considering its neighbourhood $\mathbf{V_i}$, ν_P is the calibration parameter for land use suitability, χ_P is the calibration parameter for accessibility, and θ_P is the calibration parameter for the neighbourhood effect.

The first element of the potential is cell suitability, which is considered as a binary value that is equal to 1 if the cell is suitable for a given land use and 0 otherwise, that is: $S_{i,s}$ = 1 if cell i is suitable for state s, $S_{i,s}$ = 0 otherwise. The choice for a binary variable is related to the fact that suitability and zoning were considered as one single input. It is assumed that planning regulations define suitable and unsuitable land uses for each cell, according with a predetermined zoning. Therefore, if one cell is classified as suitable for a given land use then it is assumed as being capable of attracting that land use.

The next element is accessibility. It is assessed considering travel times between cell centroids over the road network. The network is classified in hierarchical levels – e.g., main roads, secondary roads, and local roads – according with road capacity and legal speed limits. The value of accessibility for cell *i*, A_i, is assessed through the following expression:

$$A_i = 1 - \frac{f\left(T_i^*\right)}{\left\| \sum_{j \in \mathbf{C}} f\left(T_j^*\right) \right\|}, \forall i \in \mathbf{C} \tag{4.2}$$

where $f\ (T^*_i)$ is an impedance function (typically an exponential function or a power function) of an aggregate measure of travel time given by

$$T^*_i = \alpha_A \times T_{i,C} + \beta_A \times T_{i,V} + \gamma_A \times T_{i,I}, \forall i \in \mathbf{C} \tag{4.3}$$

and $T_{i,C}$ is the travel time from cell *i* to the municipality's main town, $T_{i,V}$ is the travel time from cell *i* to its civil parish (or district) main village, $T_{i,I}$ is the travel time from cell *i* to the closest industrial site located in the municipality, and α_A, β_A, and γ_A are calibration parameters. Accessibility is based on the assessment of the proximity of a cell to the urban functions available at the municipality's main town and the civil parish main village, and to the employments offered not only in urban areas but also in industrial sites.

The last element of the potential is the neighbourhood effect, $N_{i,s}$. It is considered as an aggregate value of the interactions $N_{i,s|j,r}$ between the states (or land uses) *s* and *r*, located in two neighbouring cells *i* and *j*. It is calculated through the following expression:

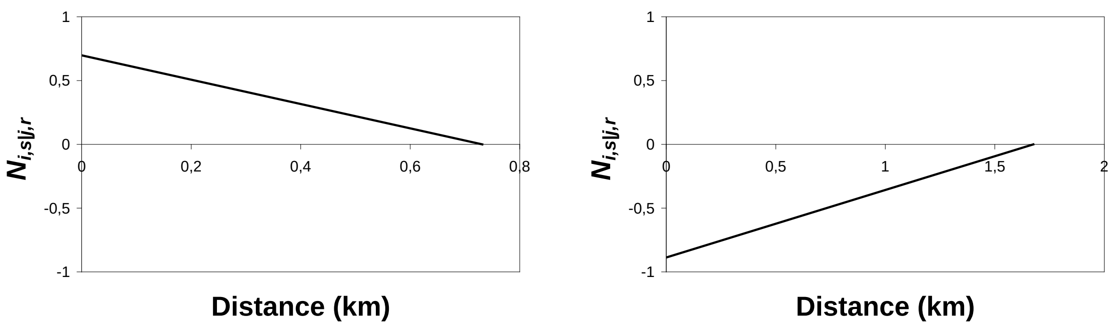
$$N_{i,s} = \sum_{j \in \mathbf{V}_i} N_{i,s|j,r}, \forall i \in \mathbf{C}, \mathbf{V_i} = \left\{ j \in \mathbf{C} : d_{ij} \le \delta \right\}, s, r \in \mathbf{S} \tag{4.4}$$

where the neighbourhood effect $N_{i,s}$ is the sum of interactions $N_{i,s|j,r}$ between state *s* in cell *i* and all the states of neighbouring cells *j* that belong to neighbourhood $\mathbf{V_i}$ considering the neighbourhood distance parameter δ (which means that attraction or repulsion will only be taken into account if the cell is located within the neighbourhood $\mathbf{V_i}$) and $d_{i,j}$ is the distance between cells i and j.

For each pair of states, these interactions $N_{i,s|j,r}$ were considered to be a linear function of the distance between the cells, as depicted in Figure 4.1. The value of the function is 1 if the interaction between the states of two cells is characterized by maximum attraction, 0 if they do not interact, and -1 if the interaction is characterized by maximum repulsion. This value is obtained as follows:

$$\begin{cases} N_{i,s|j,r} = \left(1 - \dfrac{d_{i,j}}{d_{s,r}^{\max}}\right) \times N_{s,r}^{\max}, \forall i, j \in \mathbf{C}, s, r \in \mathbf{S}; \text{ if } d_{ij} \le d_{s,r}^{\max} \\ 0; \text{ otherwise} \end{cases} \tag{4.5}$$

where $N_{s,r}^{max}$ is the maximum value for the iteration (the ordinate at the origin) between cells in state s and in state r, $d_{s,r}^{max}$ is the distance for which the interaction is zero.



(a) (b)

Figure 4.1 Two generic examples of neighbourhood effect relationships: (a) attraction, (b) repulsion

The transition potential of a cell is increased of a given value that depends on the distance between this cell and each neighbouring cell, if their states attract themselves. In contrast, the potential is decreased of a given value if their states repulse themselves. For example, according to the graphic depicted in Figure 1(a), a pair of cells classified in states that show mutual attraction, say UL and UH, would have their potentials increased of about 0.4

(prior to normalization) for a distance between the cells of 0.3 kilometres. Considering the opposite relationship, depicted in Figure 1(b), a pair of two cells classified in states that induce mutual repulsion, say XU and I, would have their potential reduced of about 0.3 (again, prior to normalization) for a distance between the cells of 1.0 kilometre. It is important to note that all the six states influence the potential of cells whatever their states are. However, only the five active states are influenced by the other states (all but R, where construction is highly restricted). This neighbourhood effect is a relationship that must be determined for each pair of states considered in the analysis of urban change. The fact that these relationships are very difficult to assess, as they depend on several interdependent factors such as land value, housing demand, public facility location, among several others, suggested the consideration of the two points that define them – $N_{s,r}^{max}$ and $d_{s,r}^{max}$ – as calibrations parameters.

Finally, the stochastic perturbation, ξ, is calculated through the following expression:

$$\xi = 1 + [-\log(\rho)]^{\sigma} \tag{4.6}$$

where ρ is a random variable uniformly distributed in the interval]0,1[and σ is a control parameter for the adjustment of the size of the perturbation (White and Engelen, 1993). This term has a highly distorted distribution so that most values are near one and much larger values occur only infrequently. The main purpose is to introduce the stochastic behaviour of agents in the transition potential.

The transition potential is calculated for every state and cell, being the final cell potential the highest value for the set of states. For comparison purposes the potential must be normalized, being its value given by:

$$P_i = \max\left\{\frac{P_{i,s}}{\left\|\sum_{s\in\mathbf{S}} P_{i,s}\right\|}\right\}, i \in \mathbf{C}; s \in \mathbf{S} \tag{4.7}$$

4.2.5 Time

The last key component of a CA model is time. Usually, the time step used in CA models is one year long. The model can use time steps of one, five, or ten years. The definition of the time step is a user decision based on the case that is being modelled. It can be connected with data availability, as the use of a time step similar to the censuses time lag frames CA dynamics with consistent census data. Also, it can be related with the rate at which a given cell changes state. This rate depends in particular of the planning environment – in highly regulated systems these changes are relatively uncommon.

4.2.6 Land use demand

The model deals with land use demand differently from classic CA models. Two irregular cells most probably have different areas, with different values for population or employment as well as for building density, supplying the land use market with different amounts of land. If demand was assessed through the number of cells that change state (as is usually the case with CA models), the sum of the newly occupied areas would not match the increase in population or employment that generated the demand. Therefore, the model allocates population or employment considering given thresholds for their densities (that are model inputs) in order to guarantee that it is able to simulate the observed value of land consumption for every state. The demand for urban land uses (UL and UH) depends on three main variables: the variation of population, the variation of household size, and the variation of building density. The increase in the number of households (single persons, single parent families) and the subsequent reduction of the average size of households implies the use of more space for housing, thus increasing land use demand. At the same time, building density tends to decrease over the years because of increasingly exigent requirements of public facilities and public space. The best way to relate these variables is through population density. The model calculates the increase of population during the reference period and distributes that population over the territory, considering thresholds of population densities. Land use demand is the surface of land that is necessary to accommodate this increase in population. The thresholds are calculated for different levels of building density, for different moments in time. Similarly, the demand for the industrial land use (I) is modelled based on employment densities, considering observed values of employment per area for the existing industrial sites.

4.3 Model performance

Models must be adjusted to the reality they are supposed to address. For that it is necessary to implement in each model at least a measure of performance or fitness of adjustment. This fitness measure is a function of the set of calibration parameters and provides an indicator of how close to reality the simulation is. Each set of calibration parameters create a unique value for the fitness function, which must be evaluated towards the same value for reality. The assessment of the model performance was made using contingency matrices for reference and simulation maps and the corresponding *kappa* index (Couto, 2003). The contingency matrix for two class maps is a matrix where each element, m_{sr}, expresses the number of cells classified in state s in the simulation map which are in state r in the reference map. Several comparison measures regarding the degree of agreement between two class maps can be extracted from a contingency matrix. The most used one is the *kappa* index, which value gets closer to one as the similarity between the two maps increases. In this model, a modified version of the *kappa* index was used to avoid the distortion that would have been produced if states that cannot take part in the urban dynamics – that is, state R (highly restricted uses) – were included in the computation. The consideration of cells in this state would be misleading by producing a large – though meaningless – agreement between simulation and reference maps. The modified *kappa* index, k_{mod}, used as fitness function was calculated as follows:

$$k_{mod} = \frac{n\sum_{i\in\mathbf{S}^*} m_{ii} - \sum_{i\in\mathbf{S}^*}\left(\sum_{j\in\mathbf{S}^*} m_{ij} \times \sum_{j\in\mathbf{S}^*} m_{ji}\right)}{n^2 - \sum_{i\in\mathbf{S}^*}\left(\sum_{j\in\mathbf{S}^*} m_{ij} \times \sum_{j\in\mathbf{S}^*} m_{ji}\right)}, \quad \mathbf{S}^* = \mathbf{S}/\{R\} \tag{4.8}$$

where n is the total number of elements in the contingency matrix and m_{ij} is a generic element of the matrix.

4.4 Model calibration

The calibration of a model can be achieved mainly through two different approaches (that sometimes may be combined): (1) performing a sensitivity analysis of each parameter's behaviour considering the other parameters fixed; or (2) running an optimization procedure for searching the set of calibration parameters that optimize the fitness function chosen for the model. The first approach is based on a group of procedures that take place after the simulation. Visual comparison can be used to quickly assess the similarity between modelled and reference maps. After this initial procedure, a series of sensitivity analyses can be performed to evaluate how each parameter varies when the other parameters are controlled. This calibration approach becomes difficult to apply as the number of parameters increases. The main reason for using the second approach based on an optimization approach is to ensure an extensive search for the parameters, leading to the best possible values for the parameters given the fitness function of the CA model. Recent trends show that optimization procedures are becoming more popular as tools to calibrate CA models. Feng and Liu (2012) use simulated annealing for calibrating the set of CA parameters of urban change in Chinese regions. (Kai) Cao et al. (2013) use multi-objective optimization based on different formulations of the genetic algorithm approach to calibrate CA transition rules. Just recently, (Min) Cao et al. (2015) proposed the used of the novel cuckoo algorithm to address the problem of transition rules.

In the present model the fitness function chosen to assess the quality of model results was the *kappa* value constrained to active cell states, k_{Mod}, as described before. The value of this measure should be as close to 1 as possible. This characteristic of the fitness function also suggests the optimization approach as a good method to calibrate the model.

The number of calibration parameters is considerably high: there are three accessibility parameters (α_A, β_A, and γ_A), four transition potential parameters (α_P, χ_P, ν_P and θ_P), the neighbourhood distance parameter and the 30 parameters of the neighbourhood effect relationships, totalizing 38 calibration parameters.

For the problem at hand, the optimization procedure chosen was the Particle Swarm algorithm (from now on referred to as PS). This recent optimization algorithm has been given promising results for complex optimization problems. PS has its origins in the simulation of social behaviours, in the study of the synchronized movement of bird flocks and fish schools. It is an optimization paradigm that simulates the ability of human societies to process knowledge (Kennedy, 1997). The member's movement in those groups is the result of the individual effort to maintain an optimum distance between him and his neighbours in the group (Parsopoulos and Vrahatis, 2002). The movement of the swarm takes place in a multi-dimensional solution space. The individual's successes influence their searches and those of their peers (Kennedy, 1997). Interaction in these groups enhances progress towards a solution, as particles benefit from their own knowledge and from their neighbours' knowledge. In this technique the aim is not the survival of the fittest but the joint effort of the swarm in finding the best solution.

Each particle has a memory of its past search history, usually called the cognitive component. It represents the natural tendency of individuals to return to environments where they experienced their best performance. Formally, it is the distance that the particle is from its personal best position. Each particle also knows the search results of the swarm, called the social component. It represents the tendency of individual to follow the success of other individuals. Formally, it is the distance that the particle is from its neighbours' best position (van den Bergh and Engelbrecht, 2006).

The formulation of PS is quite simple. It is based on a swarm of p particles that will fly through the search space during n iterations. The number of particles varies: it usually ranges from a few up to 60 particles (but there is no predetermined upper limit). The larger the swarm is, the better the search space is searched. Each particle has D dimensions: in this application of PS to the calibration of the CA model each calibration parameter is represented by a PS dimension. Hence, there will be 38 dimensions for each particle. The flowchart for the PS algorithm is depicted in Figure 4.2. Note that CA are an embedded process that is called as many times as the number of PS iterations multiplied by the number of particles.

Two vectors of particle data are necessary: one vector to record the particle's position and another to record its velocity (velocity represents the position change in each iteration).

Considering that the search space is D-dimensional, then the i^{th} particle of the swarm can be represented by its positional, D-dimensional vector $\boldsymbol{X_i}$

$$X_i = (x_{i1}, x_{i2}, ..., x_{iD})^T \tag{4.9}$$

and by its velocity, $\boldsymbol{D}$-dimensional vector $\boldsymbol{X_g}$

$$V_i = (v_{i1}, v_{i2}, ..., v_{iD})^T . \tag{4.10}$$

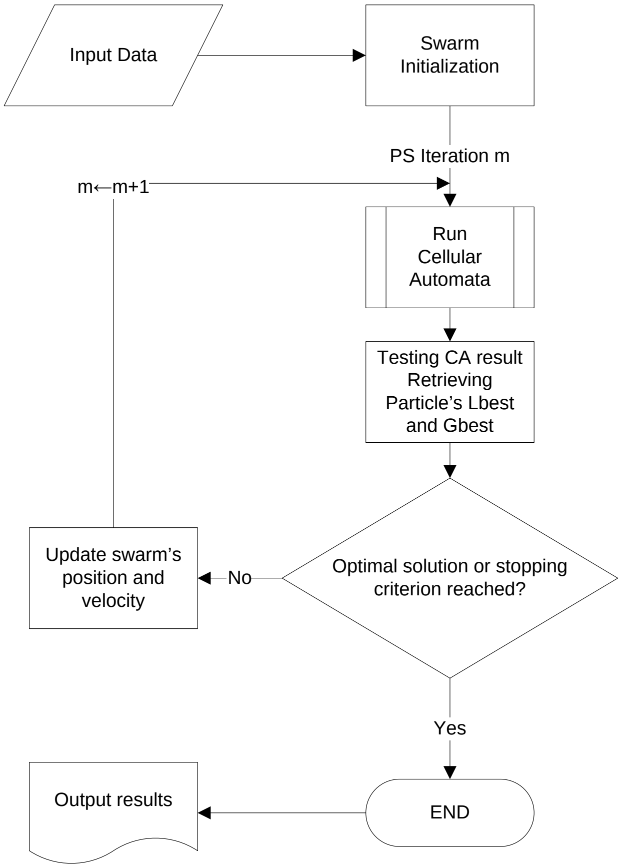


Figure 4.2 PS algorithm flowchart

The best previous position for the i^{th} particle is another D-dimensional vector $\boldsymbol{P_i}$

$$P_i = (p_{i1}, p_{i2}, ..., p_{iD})^T \tag{4.11}$$

and the global best of the swarm is represented by the last D-dimensional vector $\boldsymbol{P_g}$

$$P_g = (p_{g1}, p_{g2}, ..., p_{gD})^T \tag{4.12}$$

where g is the index for the best particle in the swarm. The velocity vector is updated according the following expression:

$$v_{i,d}^{t+1} = \omega v_{i,d}^{t} + c_1 r_1^t \left(p_{i,d}^t - x_{i,d}^t\right) + c_2 r_2^t \left(p_{g,d}^t - x_{i,d}^t\right) \quad (4.13)$$

where $v_{i,d}^t$ is the velocity of particle i on the d-dimension at iteration t, $x_{i,d}^t$ is the position of particle i on the d-dimension at iteration t, $p_{i,d}^t$ is the best individual position of particle i on the d-dimension at iteration t, $p_{g,d}^t$ is the best swarm position of best particle g on the d-dimension at iteration t, ω is the inertia factor, c_1 is the cognitive parameter, c_2 is the social parameter, r_1^t and r_2^t are random numbers uniformly distributed in [0,1]. The position vector is updated according to the following expression:

$$x_{i,d}^{t+1} = x_{i,d}^{t} + v_{i,d}^{t+1} \quad (4.14)$$

4.5 Testing the model and its calibration

The model was tested using a set of randomly generated theoretical test instances that aim to replicate plausible spatial structures. The use of theoretical test instances was oriented to exploring the performance of a CA model designed for simulating urban change phenomena in small urban areas. Another goal of using theoretical test instances regards the study of calibration procedures. The group of test instances generated for this study was produced by an algorithm that simulates spatial structures considering a set of conditions for the occupation of each cell. The main goal of this procedure is to generate spatial structures that can be considered similar to an average small municipality, both in scale and in number of cells. Cells are drawn with Voronoi polygons to represent real world census tracts, maintaining the irregularity that is a novel characteristic of this CA model. The algorithm produces land use occupation starting from a predetermined group of initial settlement centres. However, the transition of state throughout the years is based on proximity to the main network and on a probability of transition based on a predetermined neighbourhood distance. This is to say that these problems are founded on a probabilistic procedure considering a small set of conditions (neighbourhood size and road infrastructure).

These problems (of which three examples are depicted in Figure 4.3) were generated considering a maximum size of the territory (the measure of the side of the square) that ranges from 10 up to 20 km.

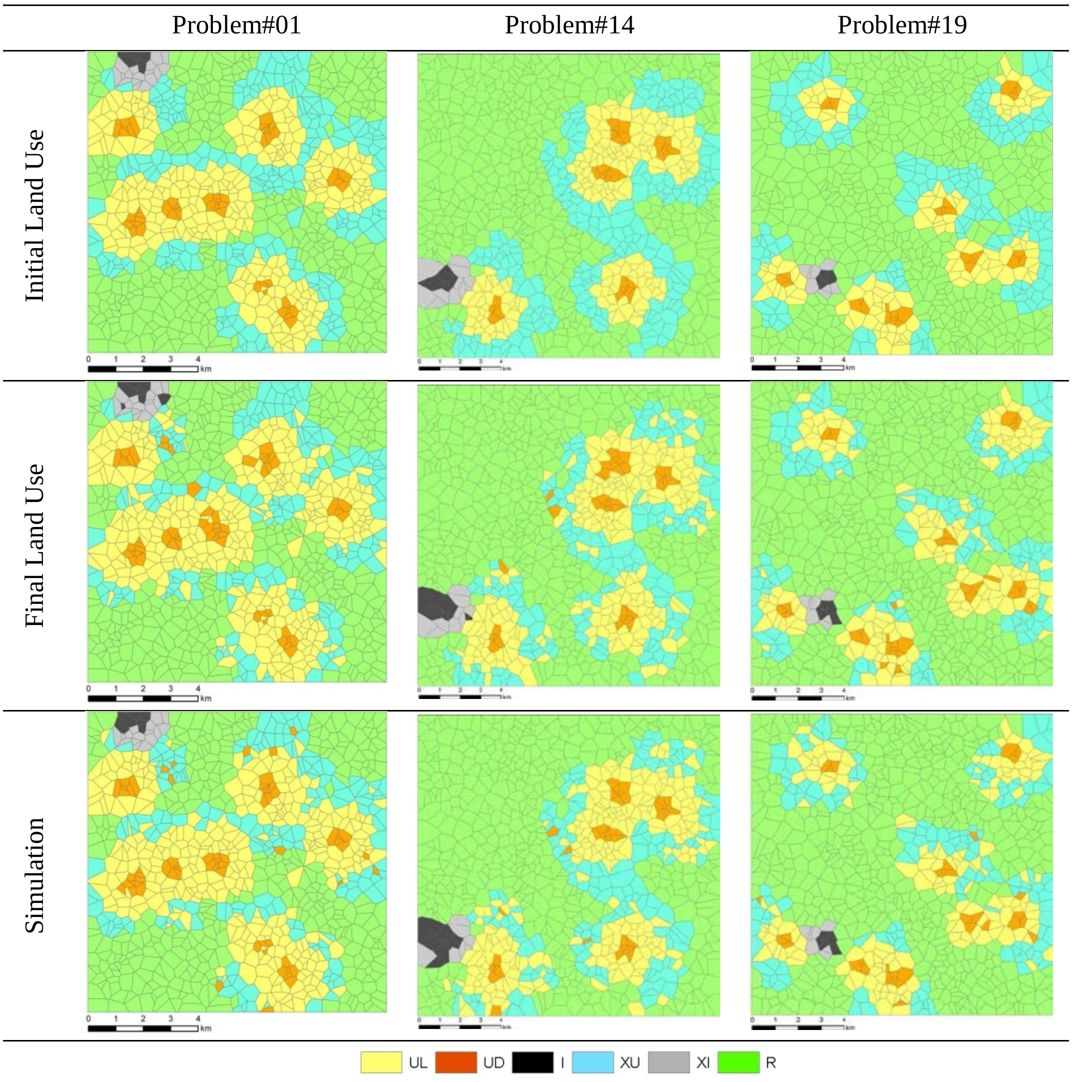


Figure 4.3 Three examples of theoretical test problems

The number of cells varies between 800 and 1200. The conjugation of these two criteria made possible the generation of different patterns, considering both the dimension and the density of cells. There are some test instances with high cell density and small size territory as well as high cell density for problems with a large maximum dimension. Average cell dimension ranges from 9.6 up to 47 hectares and territory size varies from 114 up to 380 km^2. The increase of total UL area varies from 14% up to 46%, while the same value for UH ranges from 7% and 42%. There is no direct relationship between the variations of these two sums. The change to state UH is made from cells that are both in state UH and in state XU, while the change for state UL is only possible for cells in state XU. This last state presents always a decrease on its total area, as expected. The total increase of built-up urban area (the sum of UL and UH areas) varies from 15% up to 44% with an average 27% increase.

4.6 Model results

The analyses produced for evaluating the set of test instances results focus on the performance of the model and on the assessment of model behaviour considering the characteristics of test problem. Graphic outputs for three problems are depicted in Figure 4.3.

Global *kMod* results for the entire set of problems are depicted in Figure 4.4. These results can be considered good for a simulation process: 50 percent of the problems achieved a *kMod* around 0.800 or higher and 75 percent of them exceeded 0.750. As it was explained before, *kMod* is a measure of agreement between modelled and reference maps that do not take into account inactive cell states. Figure 4.4 also presents the variation of the absolute *kappa* measure for the set of test problems. For 65 percent of the problems, the agreement exceeded 0.900 and 95 percent exceeded 0.850. These values are commonly accepted as good agreement between modelled and reference situations (Barredo et al., 2003). Overall accuracy (the proportion of correctly classified cells, that is, the sum of cells located in the main diagonal divided by the total number of cells) for the *kMod* measure also exceeded 0.850 for 75 percent of the cases. This is to say that the model showed a considerable capacity to simulate land use dynamics when dealing with theoretical small urban areas. This suggests that it can also deal properly with real world problems.

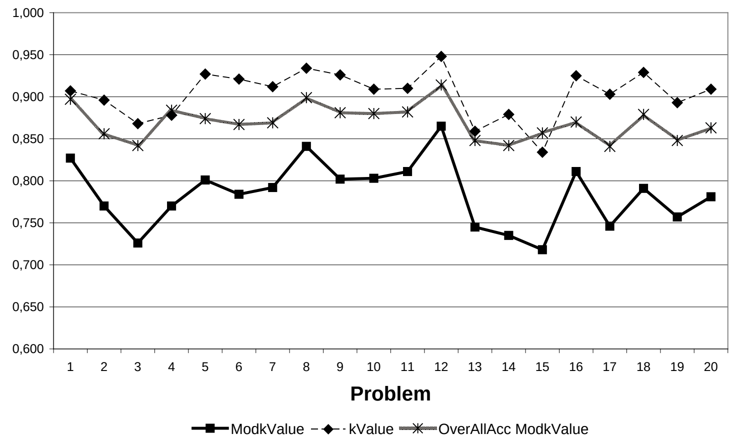
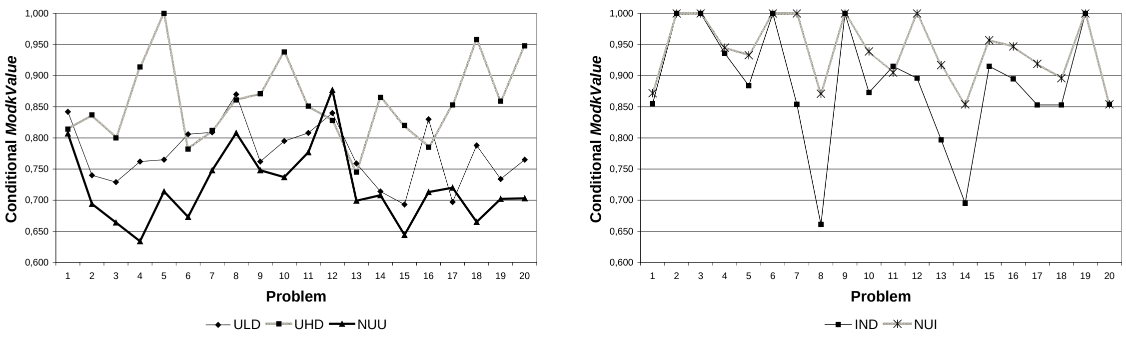


Figure 4.4 Global *kMod* and *kappa* results for the set of test problems

Another measure used to assess the performance of the simulation is the number of cells that have changed for the same state both in simulation and in reference maps. This value can be referred to as a proportion of the total number of cells that have correctly changed state in simulation. For 80 percent of the problems achieved a matching proportion higher than 35 percent with 20 percent of the entire set achieving more than 50 percent. Although these proportions present low values, they can be considered a good indicator of the model's capacity to simulate urban change phenomena. The model was unable to match a large number of state changes. However, it was able to choose cells that were close to the ones whose change was not matched. These contiguous cells have similar values for transition potential because of similar accessibility and suitability conditions.

Another measurement of good agreement can be assessed from the conditional *kMod*, the chance of agreement for each cell state within the contingency matrix. The conditional *kMod* for urban cell states (UL, UH, and XU) is depicted in Figure 4.5 (a). It is notorious a variability of the values not only along the set of problems but also within cell states.

Good results were obtained for UH cell state, with 80 percent of the problems with more than 0.800 for the conditional *kMod*. For the other cell states lower values were obtained, particularly cell state XU, for which only 35 percent of the problems achieve values of conditional *kMod* higher than 0.800. It is interesting to note that 40 percent of the problems have differences for UL and UH conditional *kMod* smaller than 0.050 which indicates a better distribution of the cells that changed state.



(a) (b)

Figure 4.5 Global conditional *kMod* results for the set of test problem

For industrial land uses the results obtained show higher values of agreement (Figure 4.5 (b)). The simplicity of the problems with regard to industrial land uses contributes to these good results. For 85 percent of the problems I conditional *kMod* exceeded 0.850 and the entire the set also exceeded this value for XI conditional *kMod*. And 25 percent of the entire set achieved total agreement for the two industrial land uses, with both values of conditional *kMod* equalizing 1, which means that there was total agreement between simulation and reality for both land uses.

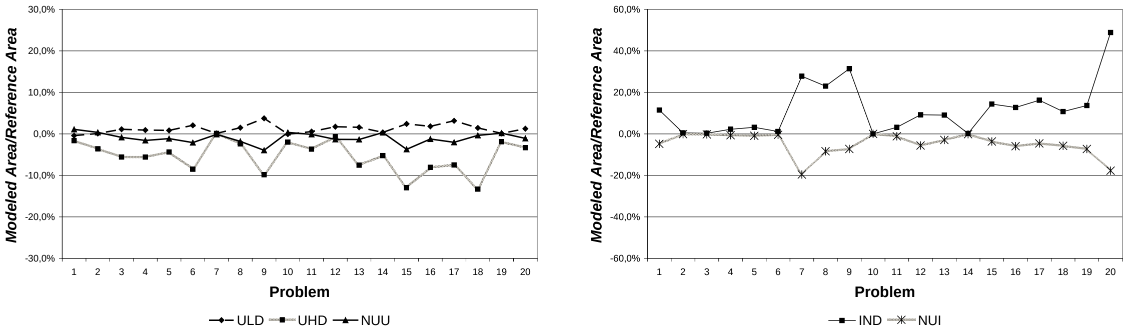
Another important parameter for model evaluation is the relationship between modelled and reference areas for each active cell state. They were compared by ratio Θ_S that is calculated through Equation 4.15 and aims to assess how different the final model outcome is from the reference map in terms of total area occupied for each lad use. It is important to assess this ratio because land use demand is considered as a function of the population and the model is oriented for the distribution of population throughout the territory.

$$\Theta_S = \frac{\left(\sum_{i \in C, s=S} \Omega_i^{Mod} - \sum_{i \in C, s=S} \Omega_i^{Ref} \right)}{\sum_{i \in C, s=S} \Omega_i^{Ref}} \times 100 \tag{4.15}$$

where Θ_S is the ratio between areas for state S, Ω_i^{Mod} is the sum of the areas of every cell i in state s=S in the simulation, and Ω_i^{Ref} is the sum of the areas of every cell i in state s=S in the reference map. The total area was determined for each cell state both for modelled and reference maps. The values for the ratio Θ_S are depicted in Figure 4.6 for every active land use for the set of test problems.

The variation of total area for UL cell state takes values between 0% and +4% with an average of +1%; for UH cell state this variation takes values between -13% and 0%, with an average -5%; for XU cell state this variation takes values between -4% and +1% with an average -1%. These values show that the model is capable of evolving to a situation similar to reality in terms of total occupied area. Differences between modelled and reference maps results from the existence of similar values of transition potential for neighbouring cells as a consequence of similar accessibility conditions and neighbourhoods. The model chooses cells near or even directly connected to cells that have changed in the reference map but not in the simulation.

The variation of total area for industrial land uses is significantly higher than the values obtained for urban land uses. The values for I cell state vary from 0% up +49% with an average of +12%. For XI cell state the variation of total area takes values between -20% and 0% with an average -5%. This behaviour for industrial land uses may be explained by an excessive simplicity of the test instances because there are few options for change in these land uses. Therefore, a small difference between modelled and reference maps (one or two cells) may result in a significant difference in the correspondent area ratios.



(a) (b)

Figure 4.6 Ratio between modelled and reference area by cell state, Θ_S

Neighbourhood distance δ varies between 2.1 and 8.0 km with an average 3.9 km. The accessibility calibration parameters show some variability in their results. For 45 percent of the problems α_A (calibration parameter for the distance to the municipality main town) was the maximum value among all the accessibility parameters. The other parameters, the calibration parameters for the distance to the civil parish β_A and to the industrial area γ_A took the maximum value among them all for 35 percent and 20 percent each, respectively. This indicates a possible trend that may evidence the importance that the distance to the main functional centre (the municipality main town) has for the formation of the accessibility measure. Regarding the transition potential calibration parameters that aim to establish the importance of the three components of transition potential – accessibility, suitability, and neighbourhood effect – the existence of a trend seems to be more evident. Both ν_P and θ_P, the calibration parameters for suitabilities and for neighbourhood effect respectively, took the maximum value among the transition potential parameters for only

15 percent and 35 percent of the problems, respectively. The majority of the problems – 55 percent – took its maximum value for the accessibility calibration parameter, χ_P. This distribution suggests the existence of a trend that gives importance to accessibility in the account of the transition potential. However, it is believed that more sophisticated measures would improve significantly the distinction between the influence that different transition potential components have in the final outcome, thus improving the quality of the simulation.

Finally, it is important to evaluate the relationship between performance and problem size in order to assess the model's ability to simulate urban change for small urban areas. Two measures were considered: the total number of cells and the proportion of cells in active states over the total number of cells. The relationship between the measure of performance of the model *kMod* and the number of cells is depicted in Figure 4.7. There is a small correlation factor R^2 of 0.138 which corresponds to a Pearson correlation ρ of -0.372. This value indicates the existence of a tenuous linear relationship between these two variables: larger values of the number of cells are related to worst values of performance. The relationship between the performance measure and the proportion of cells in active states from the total number of cells is depicted in Figure 4.8. The proportion of active cells is considered a good parameter for assessing problem size because it only considers the number of cells that participate in land use dynamics. This relationship presents a R^2 factor of 0.221 that corresponds to a Pearson correlation ρ of -0.470 indicating a significant linear correlation between these two variables.

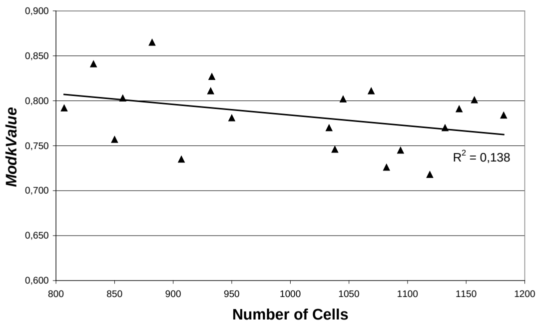


Figure 4.7 Relationship between *kMod* and number of cells

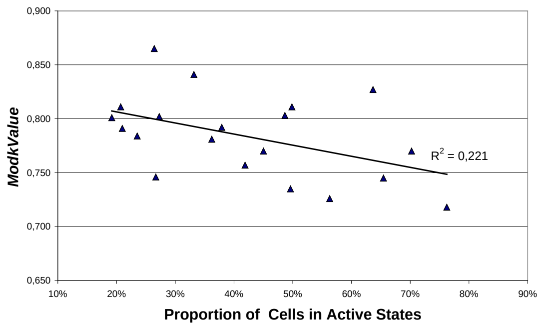


Figure 4.8 Relationship between *kMod* and the proportion of cells in active states

4.7 Conclusions

The CA model presented is oriented for the simulation of small urban areas. The results show promising possibilities of using CA for modelling urban change in small urban areas. The use of irregular cells also proved to be feasible. The model is essentially based on the assessment of a transition potential that is a function of several components, such as accessibility, land suitability, and neighbourhood relationships. Although the results for simulation agreement are closer to commonly accepted thresholds for other spatial analysis (such as remote sensing), the model showed signs that there is a significant margin to improve the assessment of the three components of the transition potential. Accessibility and land suitability must be transformed into input measures in order to lighten the calibration of the CA model. The concept of neighbourhood should be more oriented for real urban structures rather than to its mathematical concept. The assessment of neighbourhood relationships is another field that needs careful research. The use of local scale CA model is believed to produce good simulation results for urban change phenomena. The integration of CA with other models specifically designed for simulating accessibility, physical suitabilities and land use demand is believed to lead to the development of powerful modelling tools ultimately aimed to assist planning.

5 A Micro-Scale Cellular Automata Model

The problems generated by the rapid growth of urban areas raise important issues to the planning process. These issues regard not only the motives behind past evolution, but also the definition of new land-use policies capable of responding to the needs of the present and to predict and control the evolution towards a sustainable future. The complexity of the problems is such that there are no simple ways to achieve a solution nor these solutions are based on a single approach. The comprehensive nature of the planning process is simultaneously a strength, because it results from the consideration of all ingredients of the problems (physical, sociological, economic, historical, among several others), and a potential weakness, because the problems become more and more complex as the natural evolution of societies takes place, especially in the current globalized context, demanding from planners new levels of commitment and accuracy in their research and work.

For dealing with the complex problems involved in urban planning from a comprehensive standpoint, it is necessary to develop models capable of capturing the behaviour of urban systems and forecast their evolution. Cellular Automata (CA) are among the most successful urban simulation models. They have been under intensive research for the past 20 years. In a simple definition “an automaton is a processing mechanism with characteristics that change over time based on its internal characteristics, rules and external input” (Benenson and Torrens, 2004). CA are based on a discrete set of spatial units called cells that together form a cell space. Each cell takes a given (cell) state from a finite set of states. Time is considered in a discrete manner. Each cell, which works as an automaton, then operates state changes over time according to a finite set of transition rules that can be of various types (deterministic, stochastic, unconstrained, or constrained). State transition results from the application of these rules to each cell considering the neighbouring cells.

5.1 Brief literature overview

Cellular automata (CA) models are, in fact, one of the most sophisticated tools to simulate any kind of phenomena with an intrinsic spatial nature. The mathematical concept introduced by Ulam and von Neumann in the 1940s was latter introduced to geography by Waldo Tobler (1979), starting a period of intensive theoretical development that gave birth to the first applications of CA to both theoretical instances and to real world case studies (Couclelis, 1985a, 1987, White and Engelen, 1993, Batty and Xie, 1994). These theoretical foundations boosted numerous variations and improvements to geographic CA models that are now widely used for simulating increasingly more complex problems (Wu and Webster, 1998, Barredo et al., 2003, Silva and Clarke, 2005, Liu et al., 2008). The larger majority of these applications made use of regular cells derived from remote sense imagery and there is only a small group of studies that developed CA models considering irregular cellular fabrics (Semboloni, 2000, O' Sullivan, 2001a, Stevens et al., 2007).

CA models are also well suited to deal with the interactions between land use and transport, one of the most sought research topics in both spatial planning and transport studies, as well as in what is commonly known as quantitative approaches to planning or quantitative geography.

Both land use and transport make use of measures that reflect the level of performance of the transport system. Accessibility is one of this major performance outputs that can be observed and used as a mean to evaluate the performance and impacts of a transport system.

Accessibility has also been classified as a major driver of urban growth for a long time and a significant number of the models developed to simulate urban growth have different methods to incorporate accessibility and its interdependent effects with a series of other factors such as land price or household and activity location, just to name a few. A large majority of CA models also incorporate accessibility in their formulations, in particular considering it as one of the drivers that influence the system evolution through time (Santé et al., 2010). Accessibility is included as one of the components in the formulation of transition rules, which play the role of engine that drives CA evolution. However, the majority of these models consider accessibility as a cell attribute, defined as the linear distance from a cell to the nearest road (and sometimes to rail or airport) infrastructure (Clarke et al., 1997, Li and Yeh, 2000, Barredo et al., 2003, He et al., 2008, Yang et al., 2008). This somewhat simplistic approach discards the effects that infrastructure capacity and travel demand has on the performance of the transport system and their consequences on land use. Although it is not a matter for CA model to simulate the performance of transport systems, it is clearly possible to deepen the integration of land use simulation provided by the CA model with a more complex transport model that can provide a more elaborated measure of accessibility, which can include infrastructure capacity, travel demand, and multimodal systems.

The chapter presents the application of a CA model in which accessibility is a cell attribute calculated considering travel times over a real road network. This version has a similar conceptual formulation of the model presented in the last chapter but includes some innovations that were implemented with issues identified in previous runs of the model regarding accessibility and land use demand. Section 5.2 presents a description of the model with its basic formulation. Section 5.3 is dedicated to present the results for the application of the model to simulate different simple scenarios considering the construction of an important urban road for the case study of Coimbra, Portugal. Finally, section 5.4 presents the discussion over model development and the inclusion of accessibility in CA, its potential and limitations.

5.2 Model formulation

The cellular automata model presented in this chapter was designed to maintain the simplicity of the original CA concept. This section presents a brief overview about the different components of the model.

Cell is the first of the five key CA components. The model uses irregular cells that aim to simulate real-world irregular spatial partitions, such as census blocks. Neighbourhood, another key component of CA, is considered as a radial distance and it is a parameter calibrated by the model. The set of cell states, the next CA component, comprises six aggregate cell states (or land use classes): urban low density (UL) and urban high density (UH), which includes all the traditional land uses that are located inside urban areas, including public facilities and also public space; non-urbanized urban areas (XU) which are areas that could receive new urban developments; industry (I); non-urbanized industrial areas (XI); and areas where construction is highly restricted (R). Transition rules are another component of the CA model which play a key role as they represent the way the system evolves throughout the simulation. The model uses a measure of state transition potential that is used for selecting which cells will change state at each simulation step. The potential function computes a calibrated value of accessibility, land use suitability, and neighbourhood effect.

$$P^*_{i,s} = \left(\nu_P \times S_{i,s} + \chi_P \times A_i + \theta_P \times N_{i,s}\right) \times \xi, \forall i \in \mathbf{C}, s \in \mathbf{S} \tag{5.1}$$

where, for each cell i from the set of cells C, and for each state s from the set of states S, $P_{i,s}$ is the transition potential for state s of cell i, $S_{i,s}$ is the land use suitability value for state s of cell i, A_i is the accessibility value of cell i, $N_{i,s}$ is the neighbourhood effect for state s of cell i considering its neighbourhood V_i, ν_P is the calibration parameter for land use suitability, χ_P is the calibration parameter for accessibility, θ_P is the calibration parameter for the neighbourhood effect, and ξ is the stochastic parameter. Land use suitability is a binary value that has the value of 1 if a cell is suitable for a given land use and 0 otherwise, as defined by land use regulations in force. Accessibility is measured by a function of the travel time between cells (defined by their centroids) by the road network considering its hierarchical structure as follows:

$$A_i = 1 - \frac{f\left(T_i^*\right)}{\left\| \sum_{j \in \mathbf{C}} f\left(T_j^*\right) \right\|}, \forall i \in \mathbf{C} \tag{5.2}$$

where $f(T^*_i)$ is an impedance function (typically an exponential function or a power function) of an aggregate measure of travel time given by

$$T^*_i = \alpha_A \times T_{i,C} + \beta_A \times T_{i,V} + \gamma_A \times \sum_n T_{i,I_n}, \forall i \in \mathbf{C}, \forall n \in N \tag{5.3}$$

and $T_{i,C}$ is the travel time from cell i to the municipality's main town, $T_{i,V}$ is the travel time from cell i to its civil parish (or district) main village, $T_{i,In}$ is the travel time from cell i to the industrial site n located in the municipality (out of all industrial sites N), and α_A, β_A, and γ_A are calibration parameters. The neighbourhood effect simulates the spatial interaction between each pair of land uses. This interaction is modelled by a linear function that decays with distance until it is no longer observed. It has a normalized value and ranges from 0 if they not interact up to +1 if they attract each other (e.g. cell states UL and R, as depicted in Figure 5.1a) or from -1 if two land uses repulse each other (e.g. cell states I and UL, as depicted in Figure 5.1b) to 0 if they not interact. The neighbourhood effect for a given cell is the sum of all the neighbourhood interactions of this cell with all its neighbouring cells within its own neighbourhood, as defined by the neighbourhood parameter.

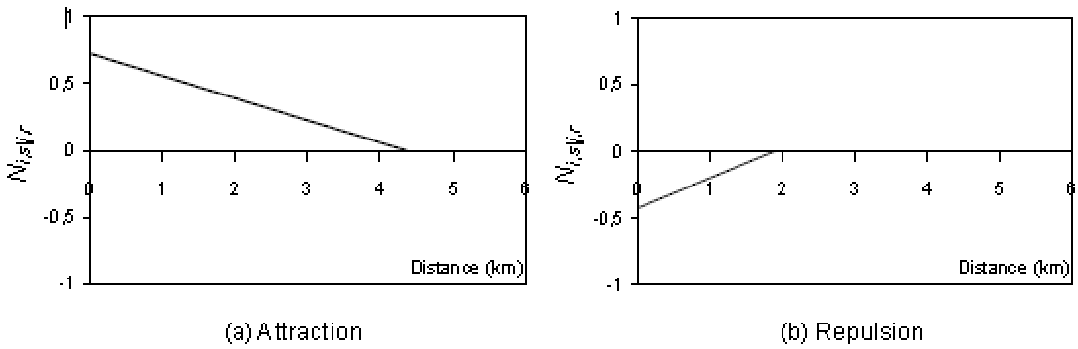


Figure 5.1 - Neighbourhood effect interactions

Land use demand is proportional to the increase of population and employment, as well as to the variation of construction density. To tackle issues of scale in newly developed cells, land demand is balanced using a standard logit model that distributes the demand considering the existent supply. The logit is applied to the value of the potential $P^{*}_{i,s}$ considering a specific control variable calibrated by the model:

$$P_{i,s} = \frac{e^{\alpha_L P^*_{i,s}}}{\sum_{i \in \mathbf{C}} e^{\alpha_L P^*_{i,s}}}, \forall i \in \mathbf{C} \tag{5.4}$$

where $P^*_{i,s}$ is the value of transition potential in cell i for state s, α_L is the logit parameter and e is the Euler number.

The assessment of model performance was made using contingency matrices and the corresponding *kappa* index (Couto, 2003). The comparison of the simulation map with the reference map through a measure of similarity is appropriate because it is oriented for the analysis of the entire territory as a distributed structure and not only as a centralized urban layout. But there are urban land uses (cell state R in the present classification) that were not considered in the changing dynamics. The consideration of the entire set of cell state for the calculation of the *kappa* index value would produce a distortion on its significance. To avoid this distortion, a modification of the *kappa* measure was considered, named k_{Mod}, accounting only the cell states that take part in the urban change dynamics. The calibration of the model was made through an optimization procedure called Particle Swarm with the goal of producing an extensive search of the set of calibration parameters that optimize the fitness function chosen for the model. This new type of optimization algorithm has given promising results for complex optimization problems. It is an optimization paradigm that simulates the movement of a group of individuals towards some goal, where the success of each individual influence its own searches and those of their peers (Kennedy and Eberhart, 1995). For an in depth reading about particle swarm see Parsopoulos and Vrahatis (2002).

5.3 Model application

The model was applied to a real world case study to simulate the impacts on urban growth of the construction of an important road ring in the urban area of Coimbra, Portugal. The main goal is to use this model application to exemplify the potential use of the model to scenario evaluation for planning purposes. The model was calibrated for a set of reference data that includes census data on demographics and employment for the years of 1991 and 2001, and considering the approved Municipal Master Plan legally in force by 2009. Two simple scenarios were designed to make a proof of concept about the possibilities of the model. There is a baseline scenario called "Baseline" which does not consider any change in the road network. A second scenario considers the construction of the road ring called 'Anel Pedrulha' (the name of both the road and the scenario). Both scenarios take into account the same values for population and employment growth rates, which were illustrative of the general trends considered in the relevant studies that supported the planning processes that led to the Master Plan.

5.3.1 The case study of Coimbra, Portugal

The model was applied to study possible scenarios of urban growth for the city of Coimbra, in the *Centro* region of Portugal (Figure 5.2).

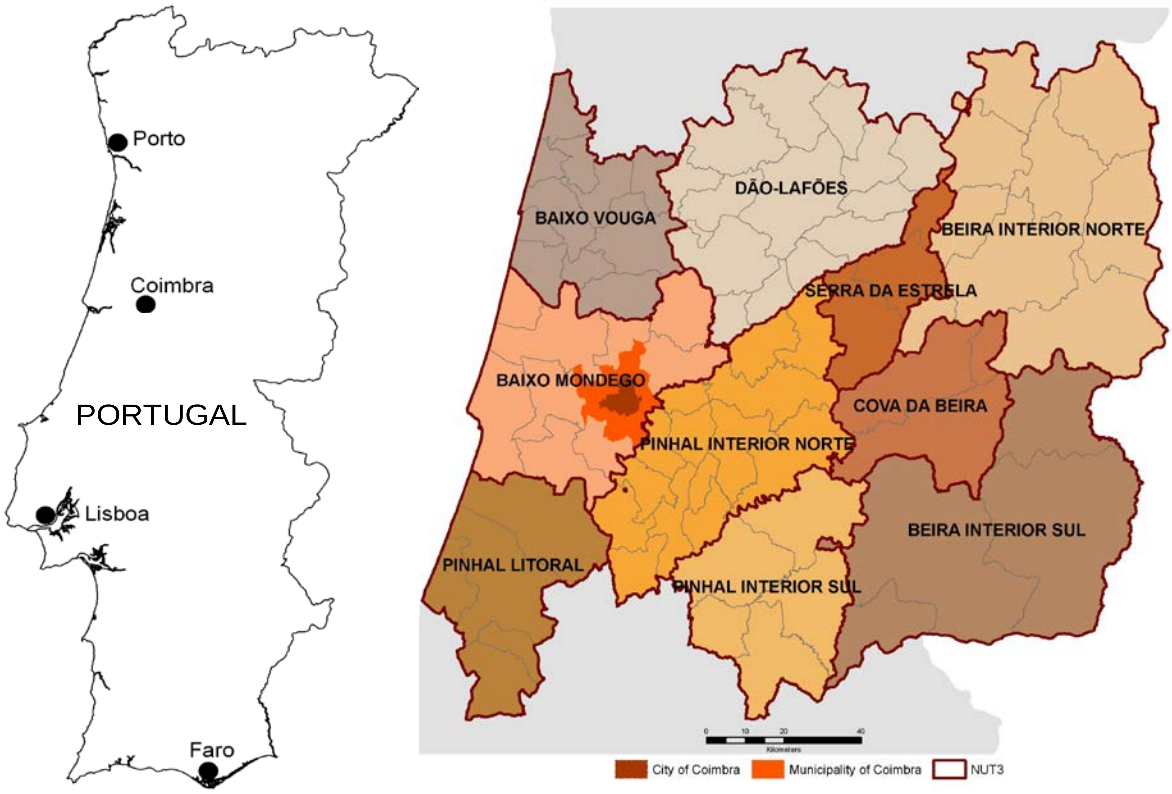


Figure 5.2 Location of the municipality of Coimbra, Portugal

Coimbra can be classified as a mid-size European city of around 100 thousand inhabitants according to the censuses (INE, 2001), which plays the role of regional capital for the central area of Portugal due to the very high concentration of public administration, healthcare and educational facilities of high hierarchical level. The city is the capital of a municipality of about 150 thousand inhabitants (INE, 2001) and heads a larger direct influence area of around 480 thousand inhabitants, the *Baixo Mondego* and *Pinhal Interior Norte* NUT3 regions. The existence of those high hierarchical level public facilities also contributes for Coimbra to exert a national influence in domains such as health, higher education and justice.

The data was formatted into a specific dataset that complies with model requirements. This dataset uses statistical data obtained from both the national censuses of 1991 and 2001, provided by the Statistics Portugal, and data from the official statistics on employment provided by the Ministry of Labour and Social Security. In addition, land use was derived from the municipal master plan in force updated considering the current occupation of the municipality. Cells were obtained from the intersection of the official census blocks from both 1991 and 2001 with the urban boundaries that are officially considered in the master plan. These cells combine urban form, derived from the urban boundaries, with reliable data from all the aforementioned statistical sources. Land use maps are depicted for the reference years of 1991 (Figure 5.3) and of 2001 (Figure 5.4).

5.3.2 Model calibration

The model was calibrated using the reference datasets for 1991 and 2001 and was able to achieve a value for k_{Mod} of 0,767 (for a *kappa* value of 0,876). The land use map for simulation results is depicted in Figure 5.5. This k_{Mod} value represents a very good adjustment of the simulation to reality, considering the standard thresholds for the use of the *kappa* statistics commonly accepted in the literature. The value of the calibration parameter for accessibility in the potential function, χ_P was of 0,665, while the parameter for neighbourhood effect, θ_P achieved the value of 0,841. The third parameter for land suitability, ν_P only reached the value of 0,278. These values illustrate the importance of both accessibility and of the interactions between different land uses for land use dynamics.

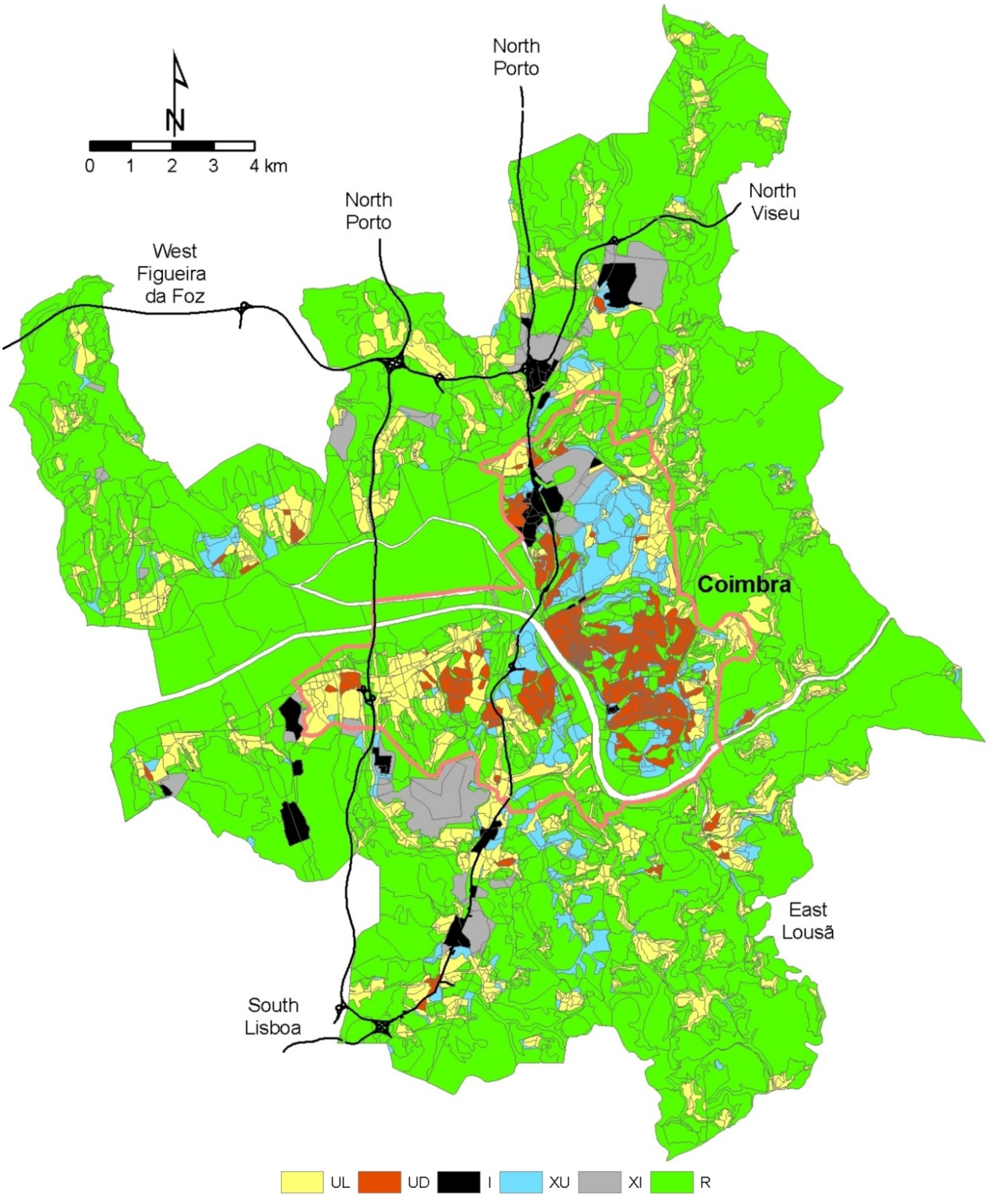


Figure 5.3 Land use maps for the municipality of Coimbra, reference data for 1991

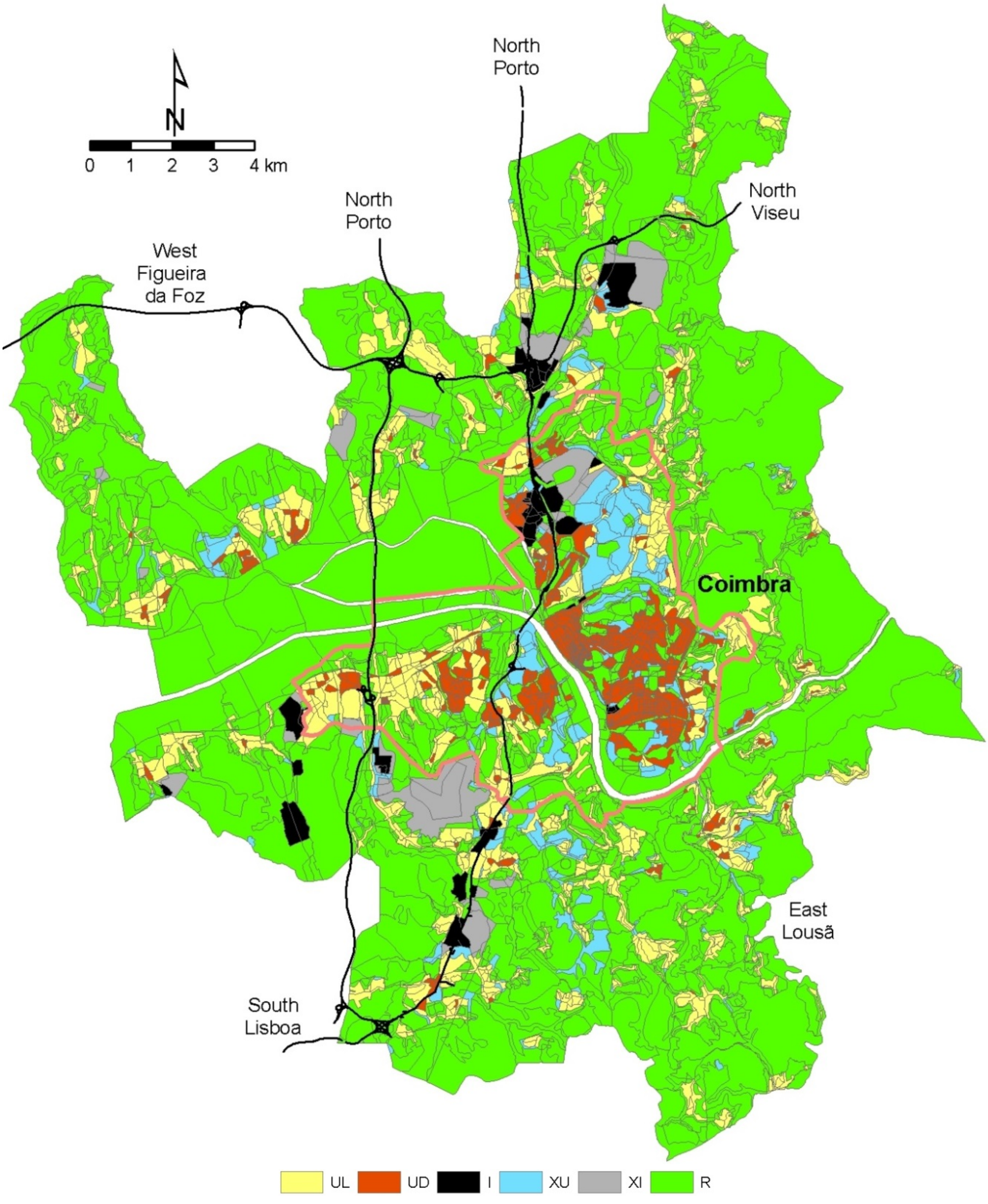


Figure 5.4 Land use maps for the municipality of Coimbra, reference data for 2001

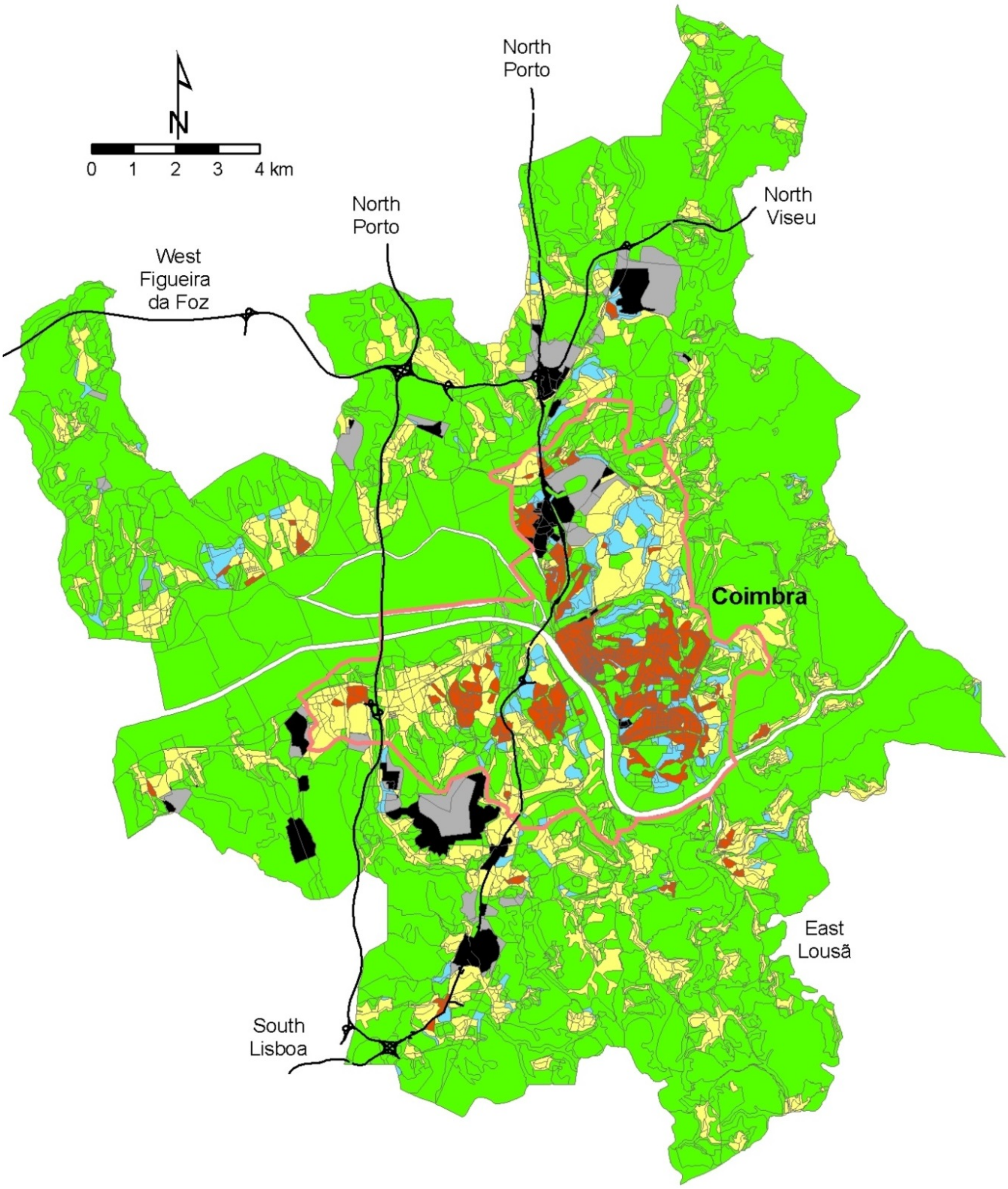


Figure 5.5 Land use maps for the municipality of Coimbra, simulation results for 2001

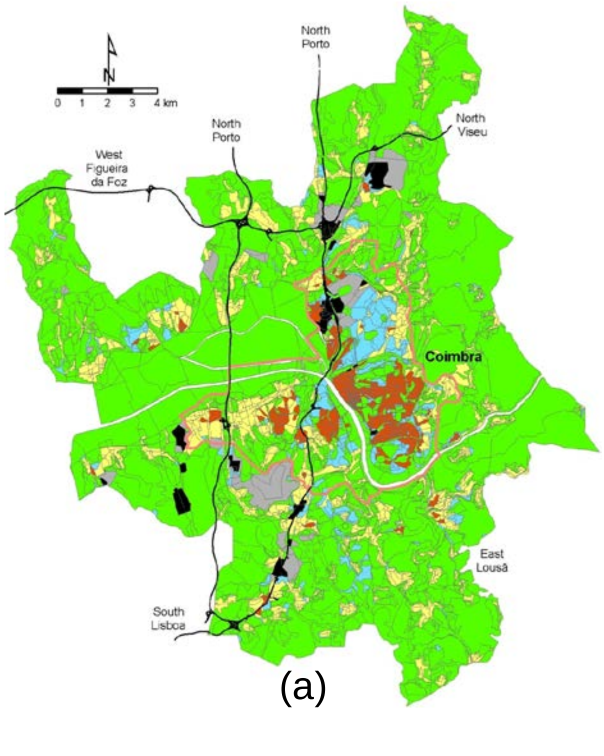


Figure 5.6 Compared land use maps for the municipality of Coimbra: (a) reference data for 1991, (b) reference data for 2001 and (c) simulation results for 2001

5.3.3 Scenario design and evaluation

Model calibration produced an optimal set of calibration parameters that can be used as the input for the application of the model as a prospective tool for land use change.

Two planning scenarios were designed to allow the evaluation of the impacts of building a new urban road ring in the northern area of the city. The two scenarios were designed considering a very simple pair of alternatives: (1) to do nothing, maintaining the same road network and therefore the same accessibility conditions; (2) to build the “Anel da Pedrulha” ring road, enhancing accessibility for many origin-destination pairs in the city OD matrix. The road maps for the two scenarios in 2021 are depicted in Figure 5.7 for the ‘Baseline’ scenario and in Figure 5.8 for the ‘Anel Pedrulha’ scenario. For both scenarios it was established the same macro conditions in terms of population and employment growth, considering two periods of ten years that will coincide with the next two inter-census periods (2001 to 2011 and then onwards to 2021), which are presented in Table 5.1. These values are in line with the forecast for population and employment evolution for the next 50 years in Portugal in the pre-crisis period, and this data was used in the studies that supported the then iteration of the municipal master plan.

Table 5.1 Macro indicators for population and employment growth for both scenarios

	2001-2011	2011-2021
Population	+3%	+2%
Employment	+2%	+2%

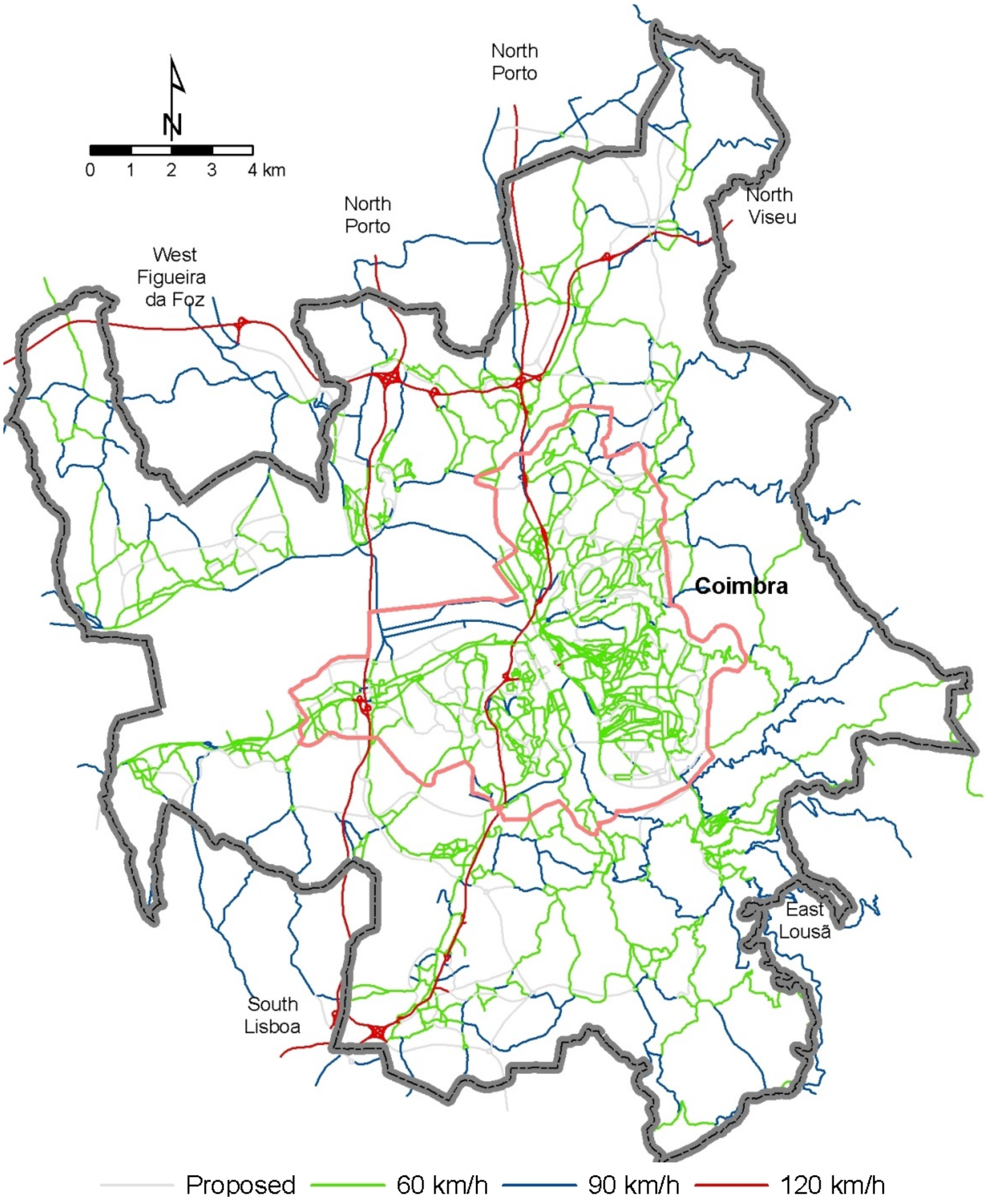


Figure 5.7 Road network of the municipality of Coimbra in 2021 for the 'Baseline' scenario

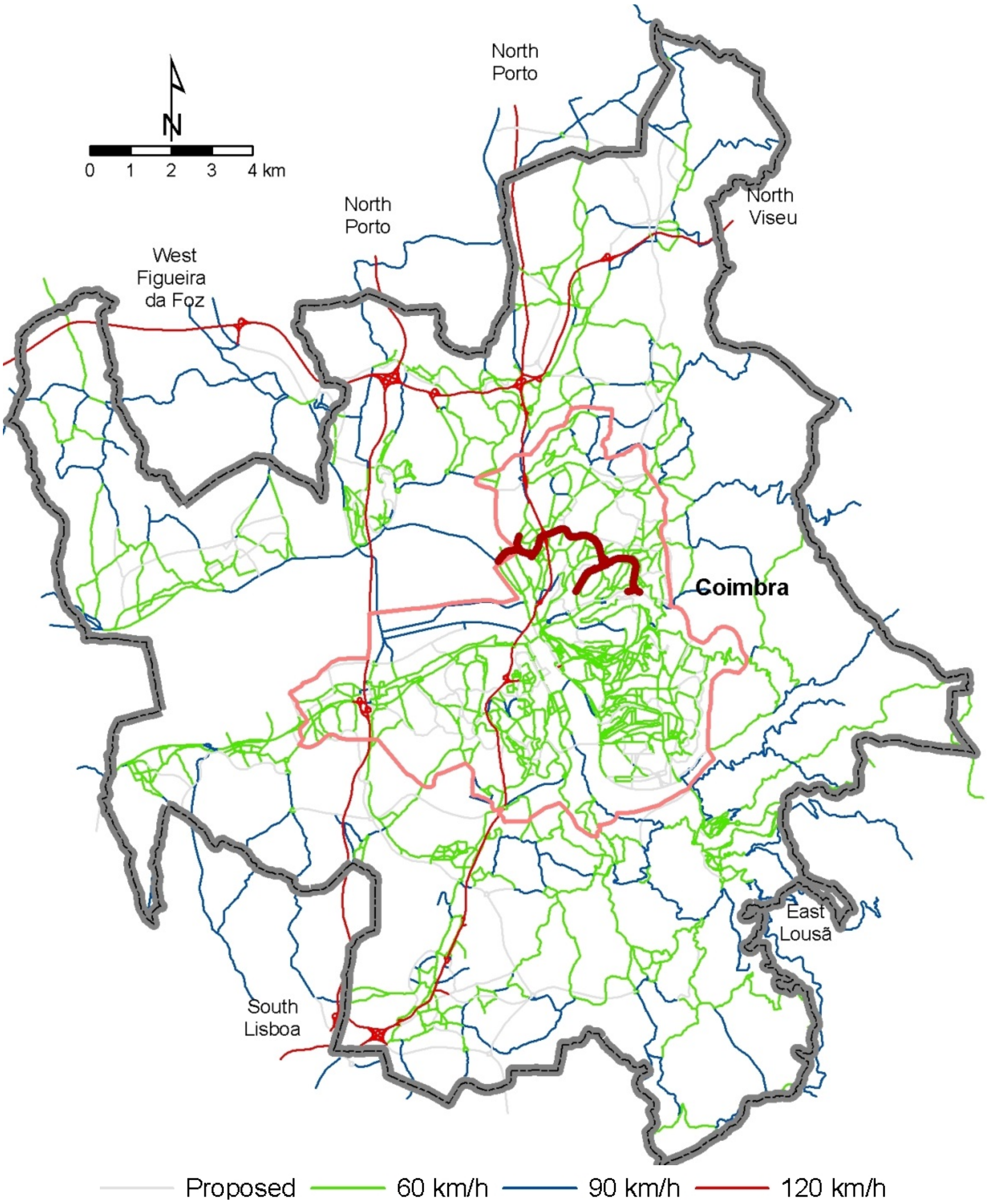


Figure 5.8 Road network of the municipality of Coimbra in 2021 for the 'Anel Pedrulha' scenario

The model used the set of calibration parameters obtained from the calibration stage with reference data to run forecast analysis for both scenarios. It is possible to see the land use maps for the ‘Baseline’ scenario for the two ten years period, with a starting reference situation in 2001 Figure 5.4 and simulation results for both 2011 (Figure 5.9) and 2021 (Figure 5.10)). Likewise, land use maps for the simulation of the ‘Anel Pedrulha’ scenario for the same years and starting from the same reference moment of 2001 (Figure 5.4) are depicted in Figure 5.12 for 2011 and in Figure 5.13 for 2021.

Figure 5.11 (for the ‘Baseline’ scenario) and Figure 5.14 (for the ‘Anel Pedrulha’ scenario) illustrate the evolution of land use from 2001 (a) to 2011 (b) and then to 2021 (c).

An enhanced detail of the area directly served by the new road is depicted in Figure 5.15(a) for the ‘Baseline’ scenario and Figure 5.15(b) for the ‘Anel Pedrulha’ scenario. These results show that the impact of the construction of the road ring is significant, as the attractiveness of cells directly served by the new road is much higher in the ‘Anel Pedrulha’ scenario, with more cells changing their state to more dense urban uses. This favouring of more dense urban uses follows what is considered In the planning documents to support the expansion of the northern are of the city of Coimbra, directly served by the road ring. The impacts on the industrial land uses is less significant as they are generally located at significant distances from the road ring and usually very close to the main collector roads, thus experiencing less impacts of the increase of accessibility that benefits more the central urban area. The effects of neighbourhood interaction (higher importance than accessibility) also played a significant factor to this reduced attraction for industrial land uses in the area directly served by the road ring.

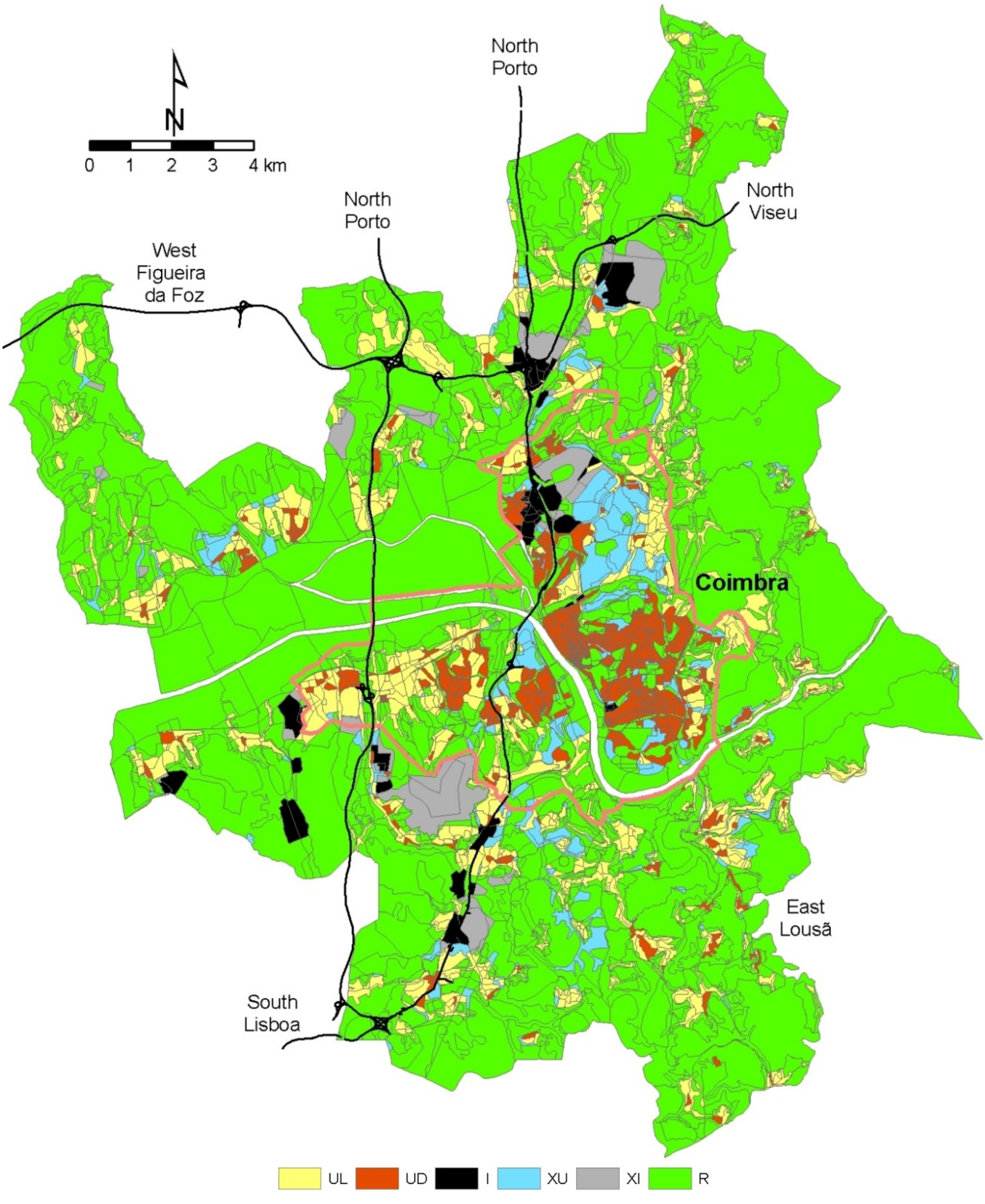


Figure 5.9 Land use maps for the municipality of Coimbra, 'Baseline' scenario for 2011

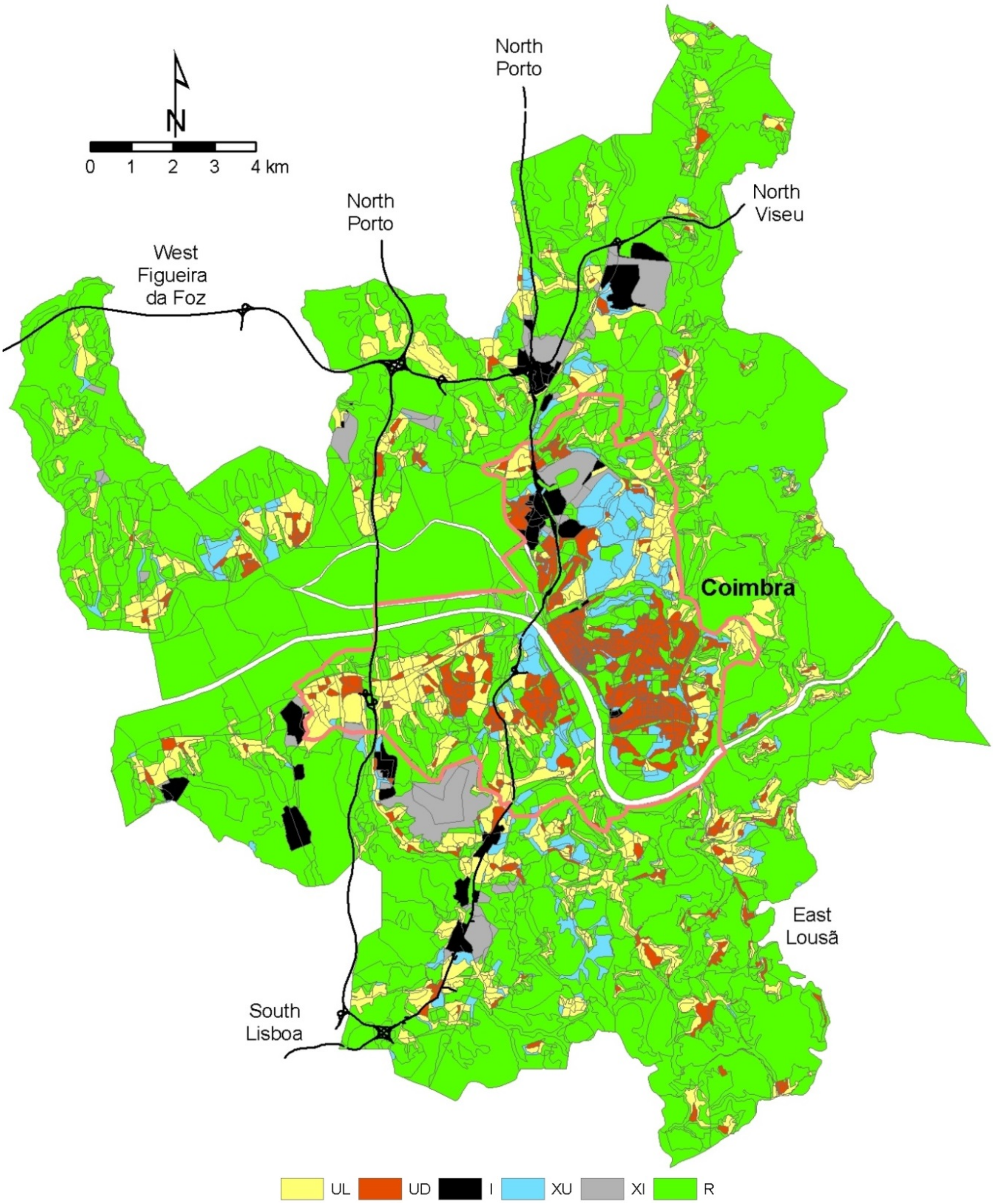


Figure 5.10 Land use maps for the municipality of Coimbra, 'Baseline' scenario for 2021

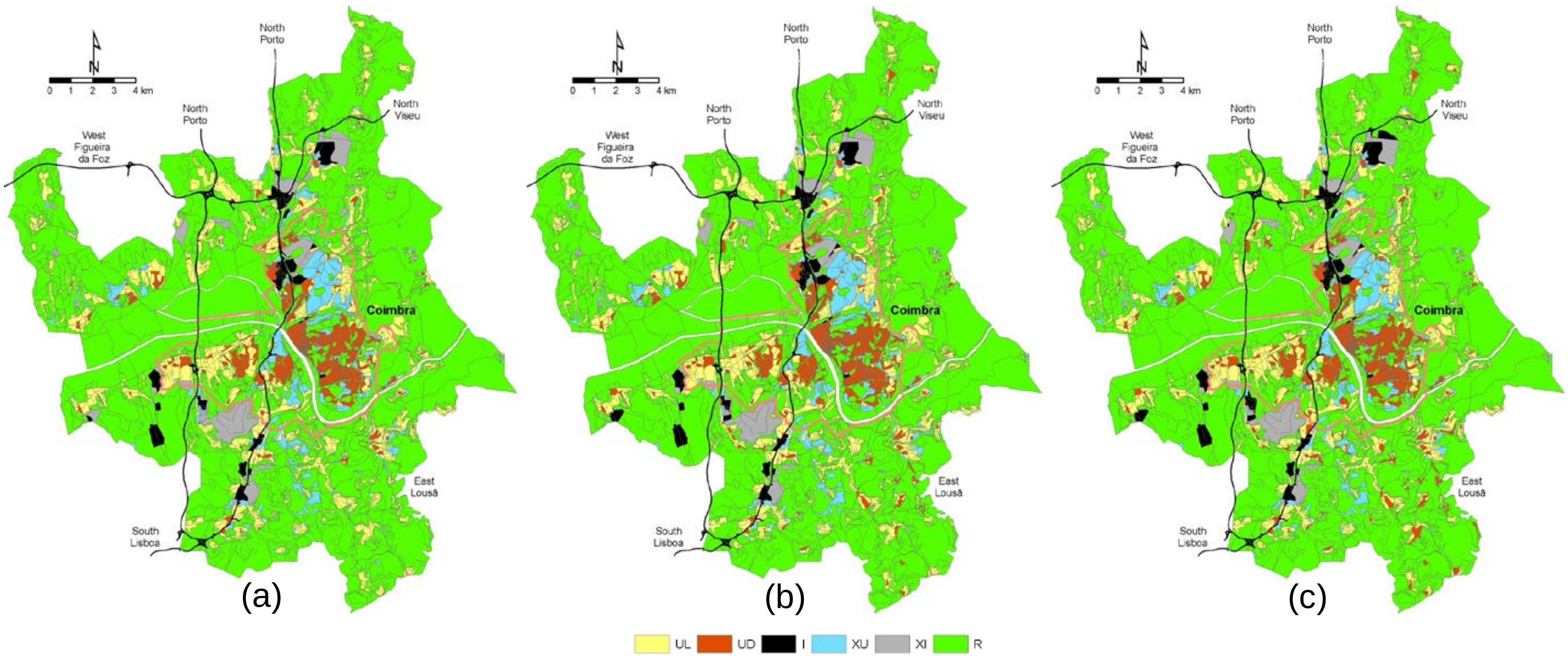


Figure 5.11 Compared land use maps for the municipality of Coimbra, 'Baseline' scenario for (a) 2001, (b) 2011 and (c) 2021

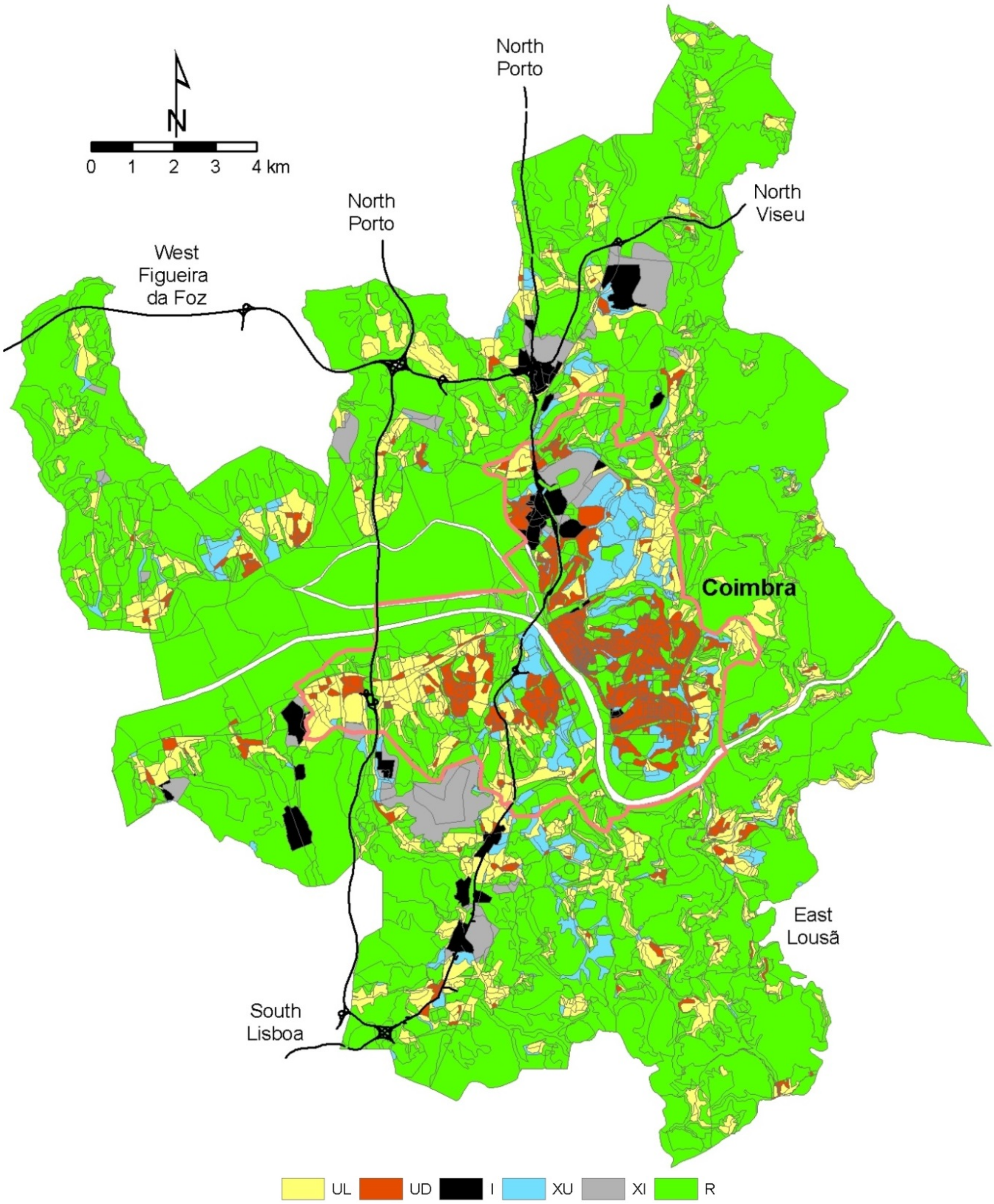


Figure 5.12 Land use maps for the municipality of Coimbra, 'Anel Pedrulha' scenario for 2011

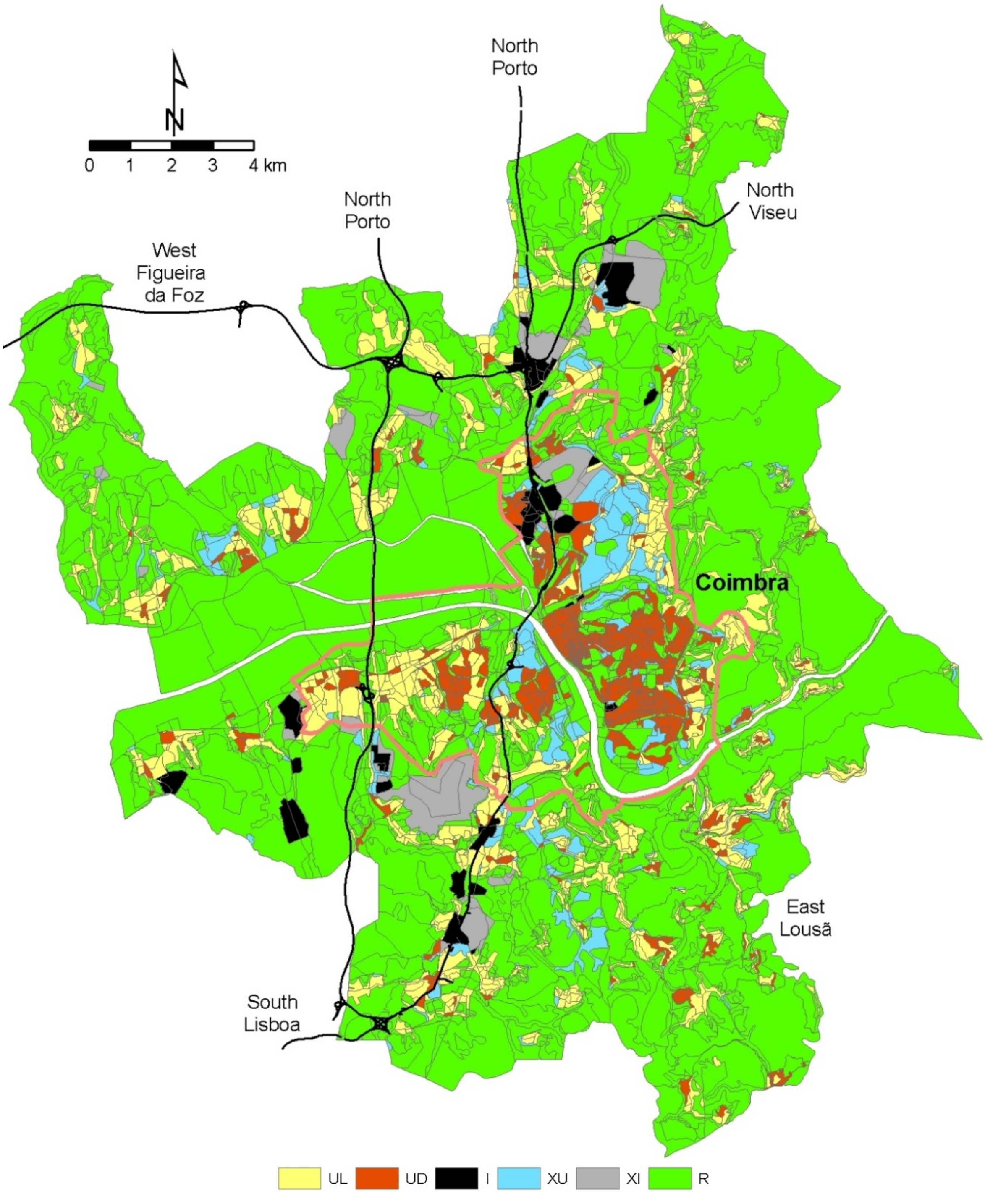


Figure 5.13 Land use maps for the municipality of Coimbra, 'Anel Pedrulha' scenario for 2021

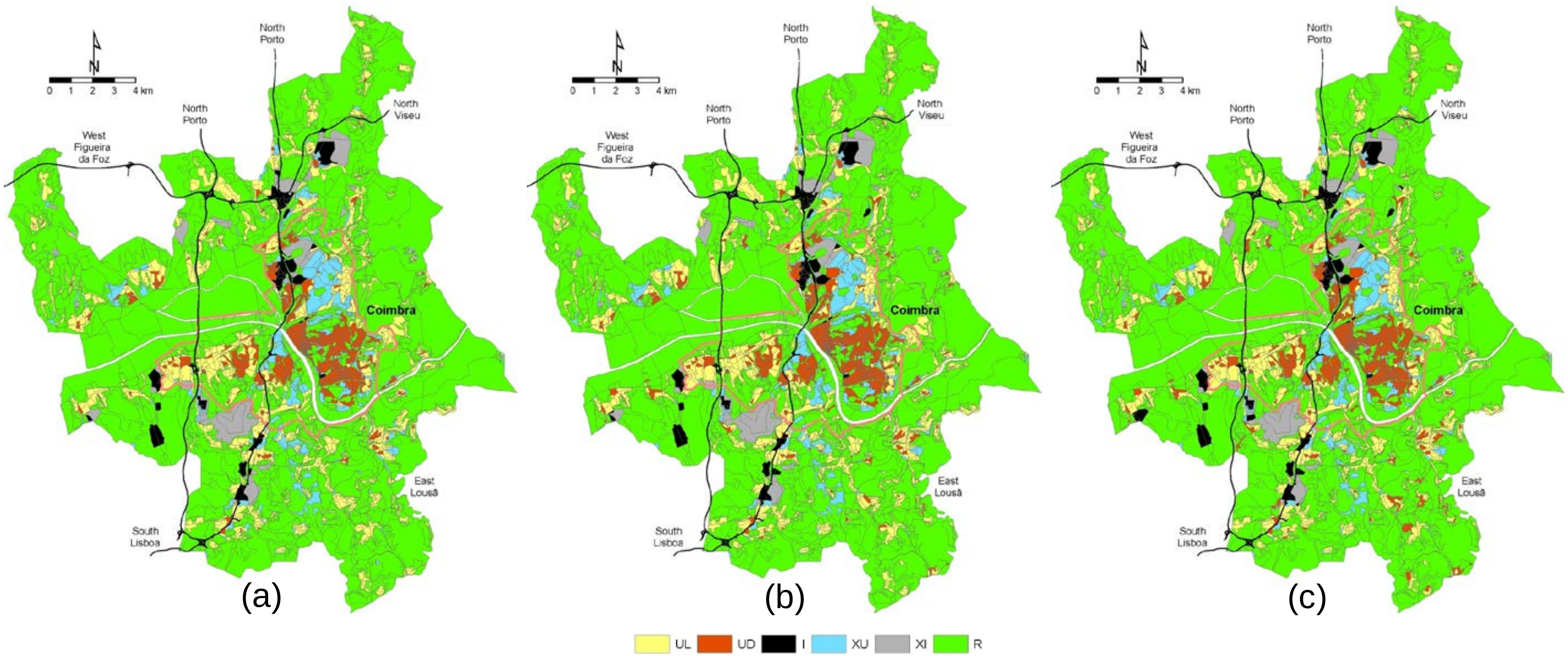


Figure 5.14 Compared land use maps for the municipality of Coimbra, 'Anel Pedrulha' scenario for (a) 2001, (b) 2011 and (c) 2021

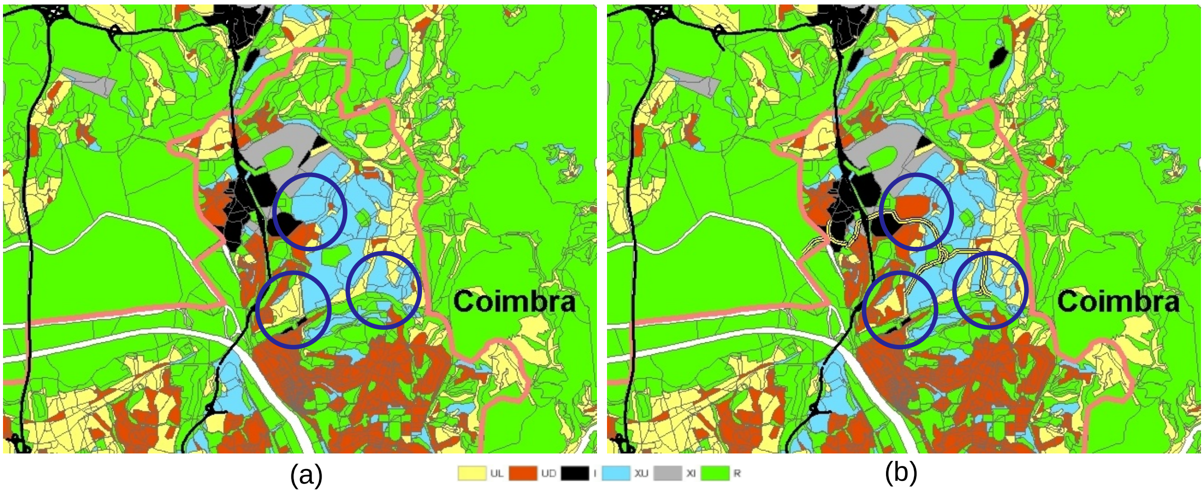


Figure 5.15 Zoom in on of the land use maps for the area directly served by the new road ring for (a) the 'Baseline' scenario and (b) the 'Anel Pedrulha' scenario

5.4 Conclusions

The model was applied to a real world case study to simulate the impacts on urban growth of the construction of an important urban road ring in the city of Coimbra, Portugal. The main goal of this application was to illustrate the possibilities of using this type of model approach to capture and forecast the complex land use and transport interactions. The scenarios designed for this application are quite simple but nonetheless very useful to exemplify what types of simple analysis are possible to be made in order to support a discussion over different planning choices. These results were presented in a workshop that focused on the application of land use and transport interaction models to both practitioners and elected officials of different backgrounds (spatial planning, transport planning) and different origins (local administration, governmental agencies). These results were useful for supporting the debate between modellers and these potential users, who engaged in a preliminary discussion over them, illustrating the potential of its use as a decision support tool.

There is a great potential in using more complex transport models to provide better accessibility indicators to CA models. First, with the use of transport models and not just spatial measures (mainly distances) accessibility is no longer just an exogenous input. It becomes a reliable measure of the performance of the transport system, by (1) taking into account effect of land use change into transport demand; and (2) providing not only an input for the CA model but also an effective lever to produce and evaluate changes in the transport system. Second, it is possible to relate land use parameters with transport parameters, testing different settings for their values in order to evaluate their interdependencies. This provides more robust tools to simulate the complexity of land use and transport interactions, enhancing the possibilities of performing policy testing to the combination of planning and transport policies and programs.

Future developments of this combined application of CA-based land use models with accessibility modelling must take into account different aspects of the transport systems, namely service levels, different scales of analysis and multimodality.

6 A Macro-Scale Cellular Automata Model

Cellular Automata (CA) models are among the most popular models for simulating spatial change and they have been developed and applied intensively during the past two decades. Two main features made CA interesting for urban studies, ever since they were introduced by Waldo Tobler in the late 1970s (Tobler, 1970): first, their inherent spatiality which suits the simulation of a wide range of geographic phenomena; second, the possibility of simulating complex patterns of, for example, land use starting from a simple conceptual framework that includes the definition of a cell space (form), a neighbourhood (interaction), and a finite set of transition rules (behaviours) applied to a finite set of cell states (land uses). This conjugation of form and function make CA models suitable for capturing the contribution of different phenomena to the complex processes of urban change.

These models are commonly used to simulate land use change at a regional or metropolitan level considering land use dynamics at a local level (Barredo and Demicheli, 2003, Silva and Clarke, 2005). They usually consider small regular cells of up to around 500 meters but they are using increasingly smaller cells, making use of the high resolution of today's remotely sensed images to capture many interactions that occur at a very coarse scale. Regular cells are used at the local scale (pixels) and at a regional scale, as aggregations of smaller cells (van Vliet et al., 2009).

This chapter addresses these issues of scale and cell form by proposing a macro-scale CA model that tries to capture aggregated land use change at a regional level, as well as to deal with population and employment dynamics. Traditional CA models operating at local scales are many times coupled with more traditional regional scale models to deal with those two main drivers (Petrov et al., 2009).

The model presented uses administrative units – municipalities or similar units, varying with the national context – as irregular cells to simulate land use change considering population and employment growth and accessibility measures at a regional scale. The use of irregular cells, regardless of the scale, is scarce in the literature (Stevens and Dragicevic, 2007, Moreno et al., 2008). It ensures a good link between form and reliable data, an approach that has been successfully applied at the local scale in the other implementations of CA models in this dissertation (see chapters 3, 4 and 5).

Scale has been debated over the years in urban modelling. The evolution of computation allowed researchers to downscale from the typical large scale models of the 1950s and 1960s to the high resolution models of current times. The debate over modelling scale has an important reference in the famous *Requiem for large-scale models* (Lee, 1973), and continued over the years, with a new moment in the mid-1990s when again the issue was brought to the agenda (Klosterman, 1994, Lee, 1994). More recently, there is again a new interest on scale, focusing also on CA models (Ménard and Marceau, 2005, Benenson, 2007, White, 2007, Briassoulis, 2008, Verburg et al., 2008).

This chapter presents a macro-scale CA model that simulates aggregate land use change at the municipal (or any other equivalent administrative unit) level. The main goal is to capture and to simulate the spatial interactions that take place at the regional/metropolitan scale where municipalities are the main decision units regarding spatial planning and urban land provision. Section 6.1 presents the model and its formulation and requirements, while section 6.2 is dedicated to its application to the Metropolitan area of Barcelona, Spain. Finally, section 6.3 is dedicated to discussion and conclusions.

6.1 Model formulation

CA models make use of five main components to simulate land use change. The model uses a partition of space, the cells, which can be classified by a finite set of states. The cell states can change according to a set of conditions, being a representation of aggregated land uses. The transition rules, which incorporate spatial interaction between states within a given spatial extent, are the engine of system evolution. The neighbourhood within which that interaction occurs, which represents the contribution of spatial interaction following Tobler's first law (Tobler, 1970). Finally, time during which the system evolves. These components are in concept scale-free, opening the possibility of applying the concept to large scale areas with large scale cells.

This macro-scale CA model is meant to keep the simplicity of the original concept of CA to remain classified as one. The model uses municipalities (or equivalent administrative units) as cells. Municipalities have the legal competence of producing land use planning for their territory, implementing public policies to promote it and to place it as an attractive location for residential and employment location.

Cell states are classified into a finite set of artificial land area classes, accounted as a percentage of the total cell area. This represents the amount of land that is somehow produced for future urbanisation, regardless of the use. It can be considered a key attribute in planning due to its irreversible nature (hardly an urbanised plot of land will be re-naturalised) and its environmental impacts, namely on the reduction of pervasive soils (and increase of water run-off flows and flood risk) and on the generation of many other externalities by the newly built stock. In order to determine land use demand at a municipal scale, the model uses the aggregate amount of land use available for urbanization, which can be related with the drivers for land demand, population and employment variation and accessibility conditions. This aggregate land use area can be determined by land use (residential, industrial, and commercial) if there is available data. Due to data availability the model uses aggregated land uses.

Spatial interaction is the other important component of CA, as it simulates Tobler's first law. These interactions take place within a variable neighbourhood which distance value is determined by the model, being a calibration parameter. The value of the neighbourhood depends on how attractive a given municipality is when all the other municipalities are concerned. This extent of attraction varies from case to case, hence its calibration by the model. The large dimension of the cells and the highly aggregate nature of the spatial interactions at stake in the model suggest the consideration of two different levels of neighbourhood. First, an overall neighbourhood δ, within which the maximum value of spatial interaction is considered. Second, an inner neighbourhood δ_i, which represents an increase of the interaction between cells within that distance.

Transition rules, the other major component of CA that aim to simulate the dynamics of the system, intend to simulate spatial interaction based on a functional transition potential that depends on the population, the employment, and a function of distance over the road network, calculated as follows:

$$V_i = \sum_j a \frac{\alpha_P P_i E_j}{d_{i,j}^{\beta}}, \forall i \in C, j \in C \tag{6.1}$$

where, for each cell *i* from the set of cells *C*, V_i is the transition potential for cell *i*, P_i is the number of residents in cell *i*, E_j is the number of registered employees in cell *j*, $d_{i,j}$ is the distance between cells *i* and *j* (from the set of cells *C*) measured by the road network, α_P is a calibration parameter and β is the accessibility calibration parameter. The parameter *a* represents the effect of the positive impacts of the potential of closer neighbours, assuming the value of 2 if cell *j* is closer to cell *i* than the inner neighbourhood distance value ($d_{ij} \leq \delta_i$), the value of 1 if cell *j* is located in a distance higher than the inner neighbourhood but smaller than the overall neighbourhood ($\delta_i < d_{ij} \leq \delta$) and 0 otherwise, meaning that no spatial interaction is accounted in the potential of cell *i*. In each time step, cells are selected by the model for urbanization though a measure of its relative probability (taking into consideration all cells) regarding the transition potential value, calculated through an application of a *logit* model:

$$U_i = \frac{e^{\alpha_L V_i}}{\sum_j e^{\alpha_L V_j}}, \forall i \in C, j \in C \tag{6.2}$$

where, for each cell *i* from the set of cells *C*, U_i is the relative probability value of cell *i*, V_i is the transition potential for cell *i*, and α_L is the calibration parameter of the *logit* model. In each time step the model will allocate a given amount of population (residents) and employments according to the relative level of attractiveness given by the utility value of the cell at that time.

The model is calibrated through an optimization procedure based on the particle swarm (PS) algorithm, which roots are in the simulation of social behaviours, in the study of the synchronized movement of bird flocks and fish schools. For further details on the theoretical grounds see (Kennedy, 1997), and on the mathematical formulation of the PS algorithm see Parsopoulos and Vrahatis (2002). This algorithm is suitable for dealing with a high number of dimensions (the model's calibration parameters) because it has a simple formulation which ensures that the complex interdependences between the parameters are taken into account in the calibration process. The objective function is to maximize a fitness measure of the quality of the simulation, assessed through the comparison of simulation and reference maps. The measure chosen is the *kappa* index calculated from the contingency matrix that compares the classification of cells in both the simulation and on the reference maps (Couto, 2003). PS makes use of a swarm of *p* particles (from a few to traditionally up to 120, but with no upper limit), each one representing a point of the space of solutions for the CA model. Each particle has *D* dimensions: in the CA model each calibration parameter is represented by a PS dimension. This swarm will fly through the space of solutions during *n* iterations. The algorithm retains the position and the velocity of each particle in each iteration, calculating their new values considering the group leader position and their individual best position. The CA model is an embedded process that is called as many times as the number of PS iterations multiplied by the number of particles. The flowchart for the PS algorithm with the CA model is depicted in Figure 6.1.

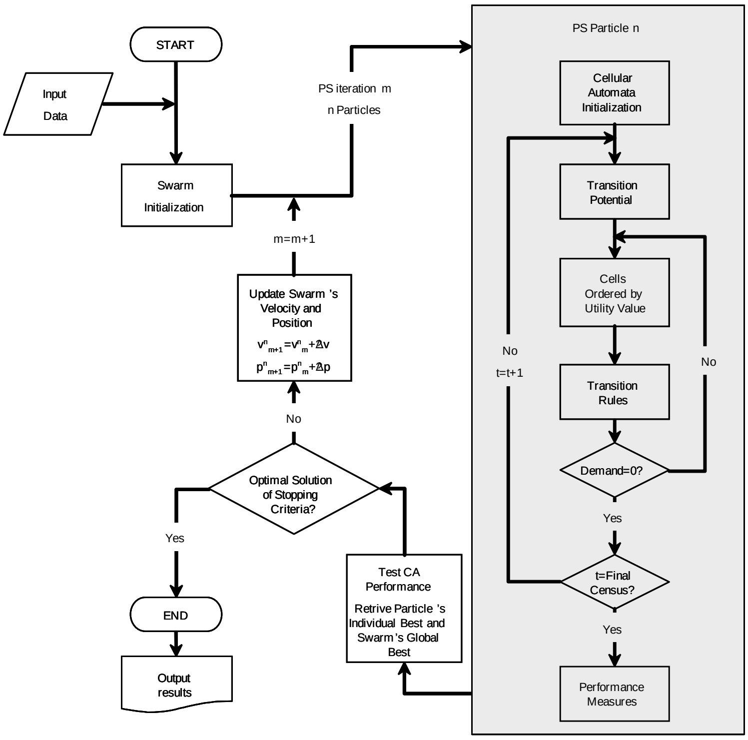


Figure 6.1 CA model (grey) and PS algorithm flowchart

6.2 Application to the Metropolitan Area of Barcelona

6.2.1 Case Study: the Metropolitan Area of Barcelona, Spain

The Metropolitan Area of Barcelona (MAB), the large city region spearheaded by the city of Barcelona, is the metropolitan capital of Catalonia, one of the autonomic regions of Spain. MAB is composed by 164 municipalities which vary considerably in area, population, and employment. Municipalities in Catalonia hold legal powers for spatial planning but have different technical and financial resources to develop and implement local policies due to the great variation in size and population. Figure 6.2 depicts MAB and shows how municipalities vary in size, with population varying from few hundreds of inhabitants up to almost a million, the same happening with employment. The city of Barcelona holds the most significant share of prominence out of a very complex set of mid-size and small urban systems which group urban areas and their hinterlands with their own functional relationships. These subsystems are served by independent infrastructure systems, but there is already a significant level of policy integration in some sectors, namely the metropolitan transit system. MAB has an area approximately 3251 square kilometres of area. In 1991, the total area of artificialized land as approximately 594 square kilometres having evolved to about 653 square kilometres in 2001, an increase of 9.9%. Despite its dimension as the second Spanish city in, the overall growth from 1991 to 2001 (the period of analysis and application of the model) for the MAB was of about 73 thousand inhabitants, from around 4.299 million inhabitants in 1991 to 4.372 million inhabitants in 2001, an overall variation of just 1.7%. Overall employment variation was much higher in value, with 383 thousand new employments in all sectors of the economy (from 1.577 million in 1991 to 1.954 million in 2001), representing a variation of 23.9%. This dataset was compiled from different sources and processed in order to suit model data requirements.

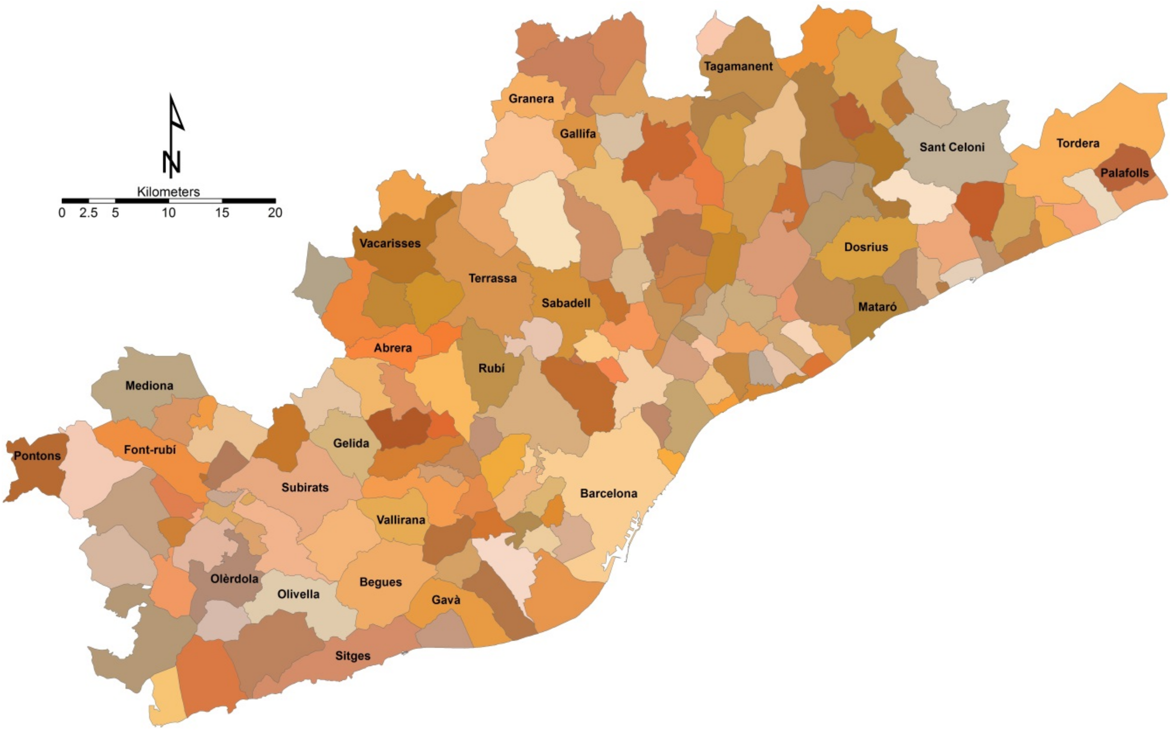


Figure 6.2 Metropolitan Area of Barcelona (MAB), with some of the municipalities identified

The model was applied to MAB in order to simulate the allocation of urbanized land over the municipalities, considering an aggregate value of population and employment density as limits for land demand. The model combines the use of demographic and socioeconomic data from the official data sources that provide data for policy design, as the Spanish Statistics Institute for population employment data, with land use data at the scale of the region provided by the European Commission's CORINE Land Cover maps. Population and employment data were collected for the census years of 1991 and 2001, while land use data was derived from CLC 1990 to match the census year of 1991 and CLC 2006 to match the census year of 2001. This simplification of the existent dataset was imposed by the difficulty of achieving a better match for both components of the dataset. Transport data was pre-processed by calculating travel times over the road network at each moment, considering road capacity (defined solely by the top speed limit) in free flow conditions. The model only considers travel on road network at this stage.

Figure 6.3 and Figure 6.4 represent the population distributions for both 1991 and 2001, respectively, using population densities measured in inhabitants per hectare of urbanised land. This measure is more representative of the population as a driver for urban change age in the model as it relates the value of the population and the value of land consumed to accommodate it. It is possible to see the great concentration of population in the main urban areas of Barcelona and neighbouring municipalities of Hospitalet de Llobregat or Badalona, the peripheral cities of Sabadell and Terrassa, Granollers or the coastal municipalities north to Barcelona up until Mataró. The comparison of the two maps show that there is a generic increase of population density (normalised by urbanised area) in and around the city of Barcelona and the municipalities of the coastal area and the main peripheral cities. This increase is mitigated by the fact that there was a significant increase of urbanised land for a small relative increase in population. The smaller municipalities of some of the off-coast areas of the metropolitan region experienced the opposite phenomena, with smaller increases in urbanised land for smaller increases of population.

Employment distribution follows a similar pattern, but with greater concentration in and around in and around the main cities of the metropolitan region, whereas the effect in the coastal municipalities is reduced when compared with population growth. Due to a significant increase of employment numbers all across MAB, it is possible to observe from Figure 6.5 and Figure 6.6 that there is an increase of employment densities (again, measured over the urbanised land area) in and around the main cities of the metropolitan region.

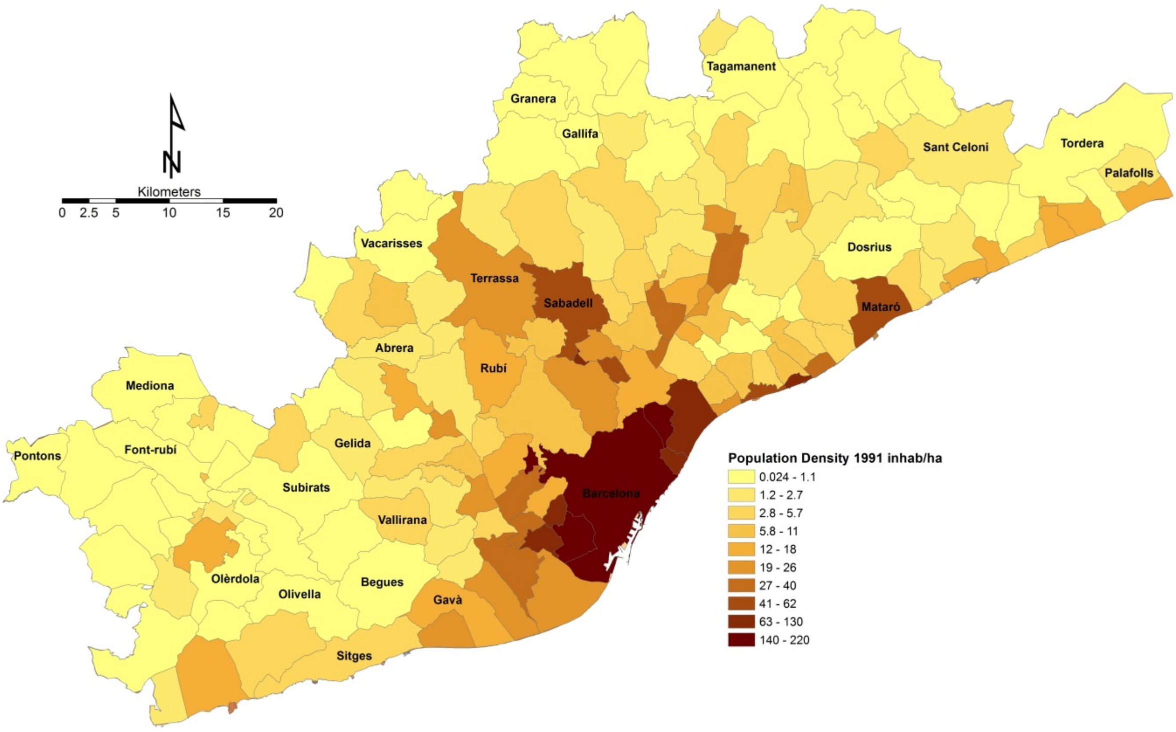


Figure 6.3 Population distribution in MAB for 1991, inhabitants per hectare

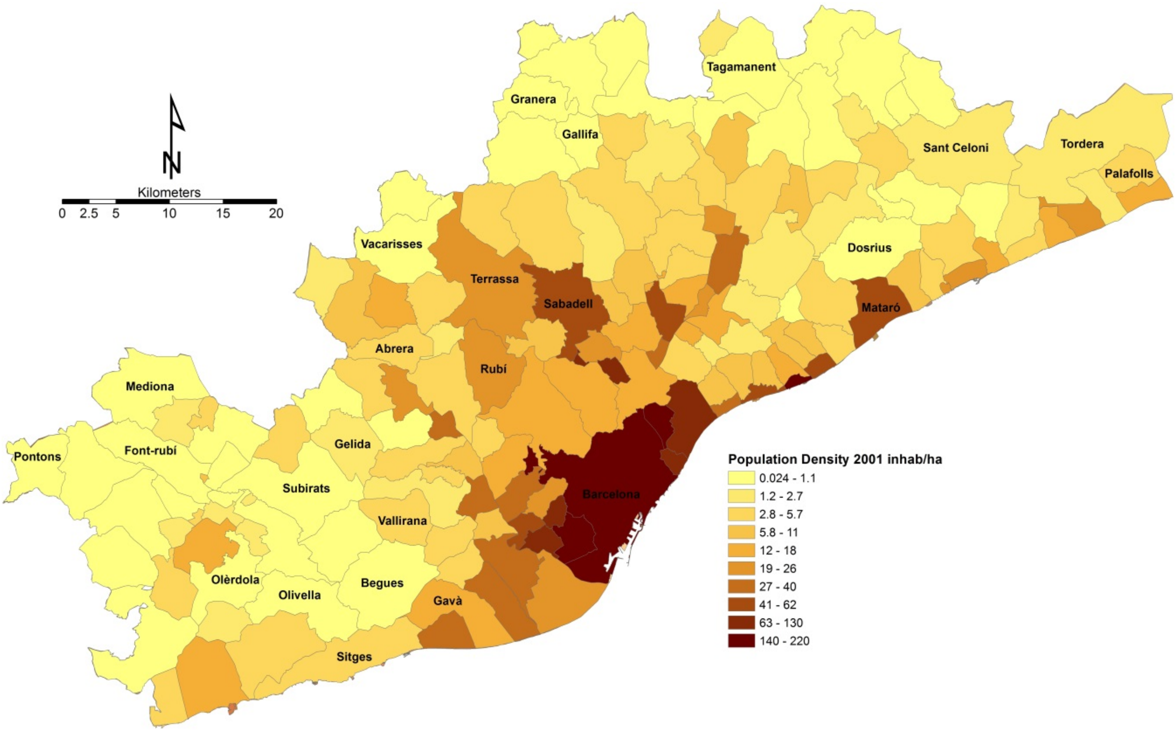


Figure 6.4 Population distribution in MAB in 2001, inhabitants per hectare

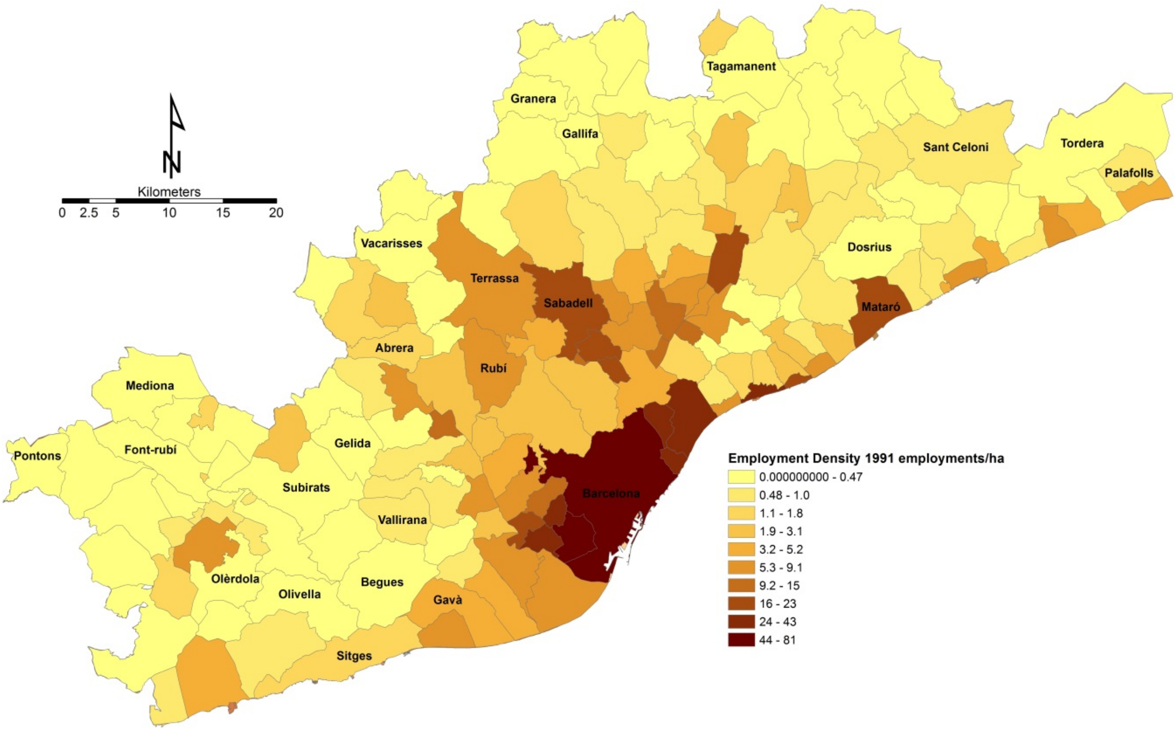


Figure 6.5 Employment distribution in MAB in 1991, employments per hectare

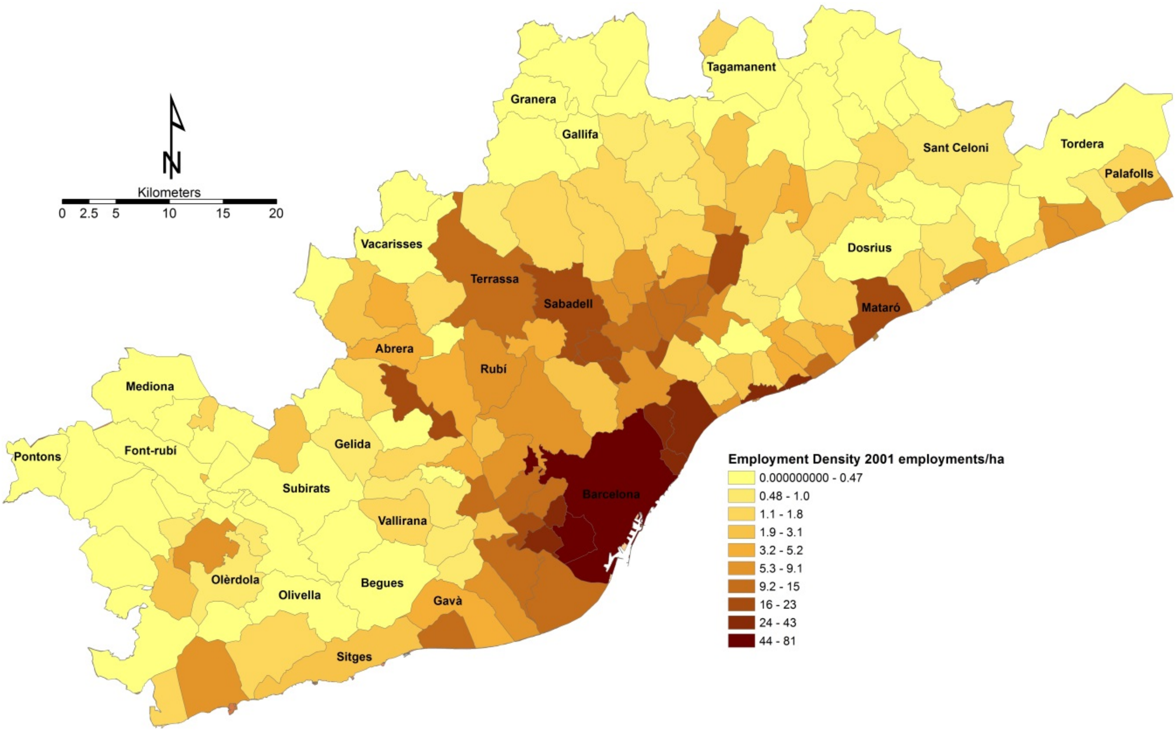


Figure 6.6 Employment distribution in MAB in 2001, employments per hectare

Figure 6.7 and Figure 6.8 depict the distribution of urbanised land over the MAB for 1991 and 2001, respectively. Urbanised land was measured in aggregate terms as the total area of land that has been urbanised and has generically lost permeability capacities, being that existent built up areas or land that potentially will be used for new urban developments. Due to the scale of analysis, the entire area classified as non-permeable land in CORINE has been considered as artificialized land, disregarding the smaller inner areas of urban green land as little gardens or the greenery that is includes in streets and similar urban features.

As for the population and employment distributions (which were normalised by using densities), it is possible to observe an increased amount of this land in and around the core of the MAB, from the city of Barcelona and neighbouring municipalities (the lower Lobregat ones or Besos) towards the Northwest, in and around Sabadell, Terrassa, Granollers, and along the coastal line, especially around its most important municipalities as Vilanova i la Geltrú, Castelldefels, or Mataró.

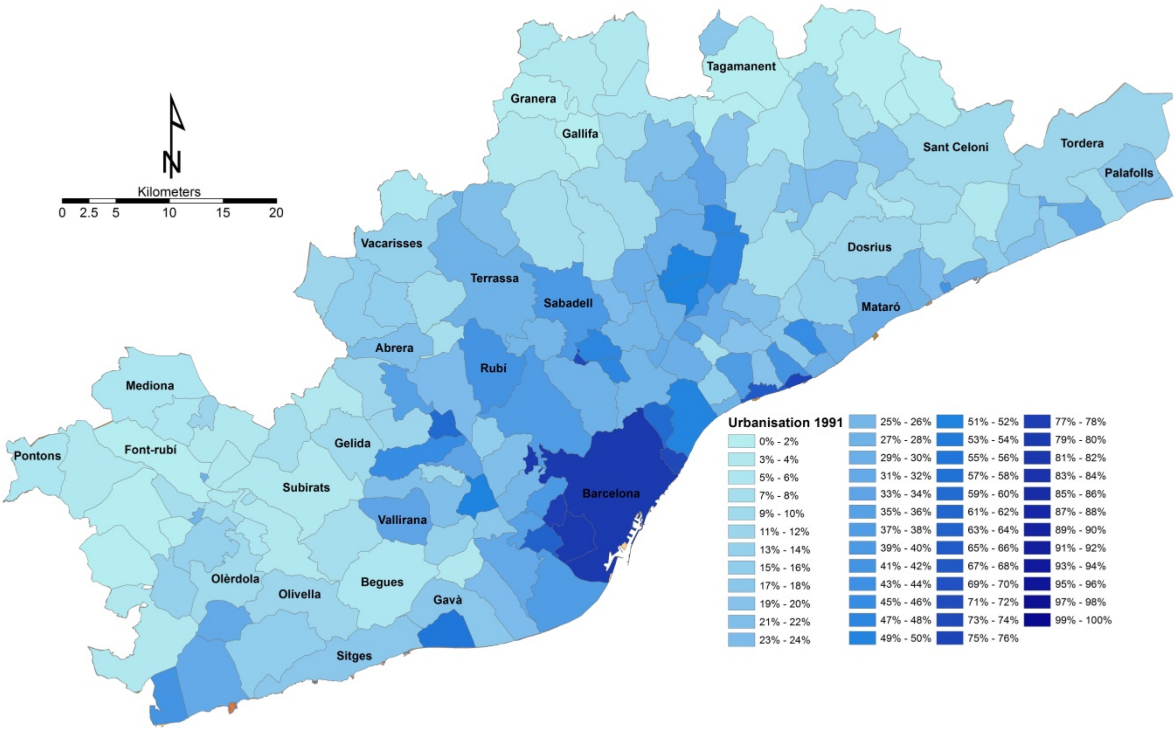


Figure 6.7 Cell states representing aggregated urbanised land for 1991, percentage of municipal area

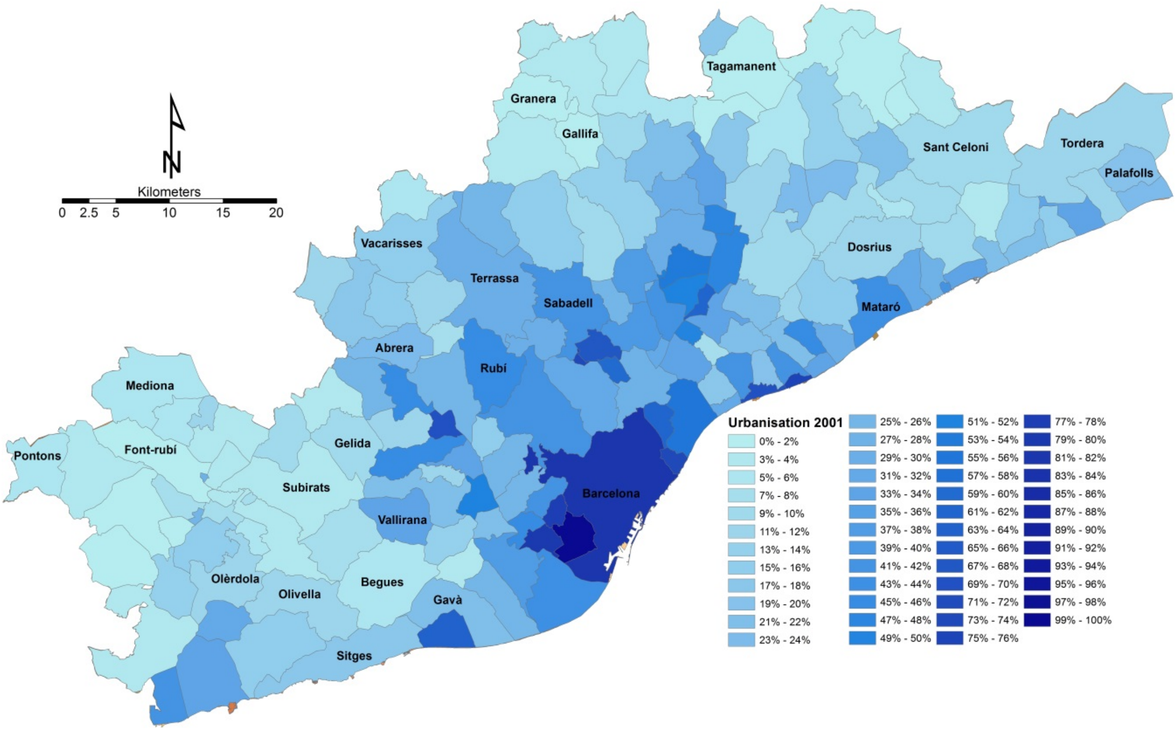


Figure 6.8 Cell states representing aggregated urbanised land for 2001, percentage of municipal area

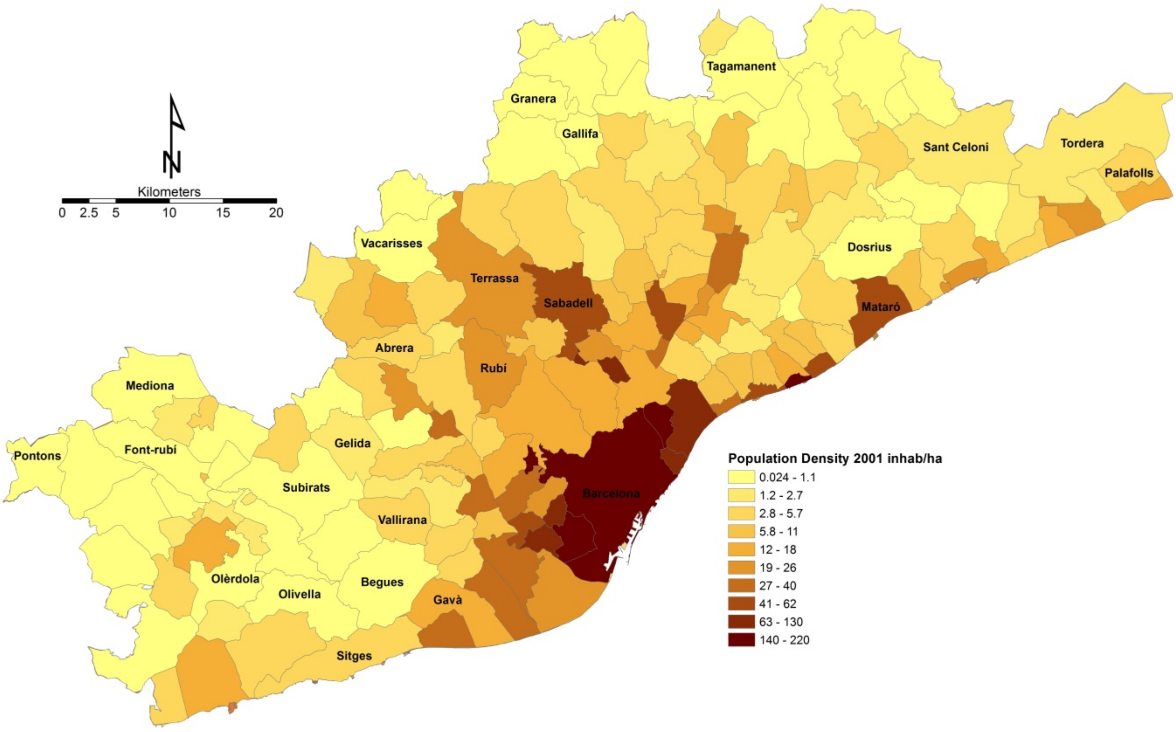


Figure 6.9 Reference map of population densities in MAB for 2001, inhabitants per hectare

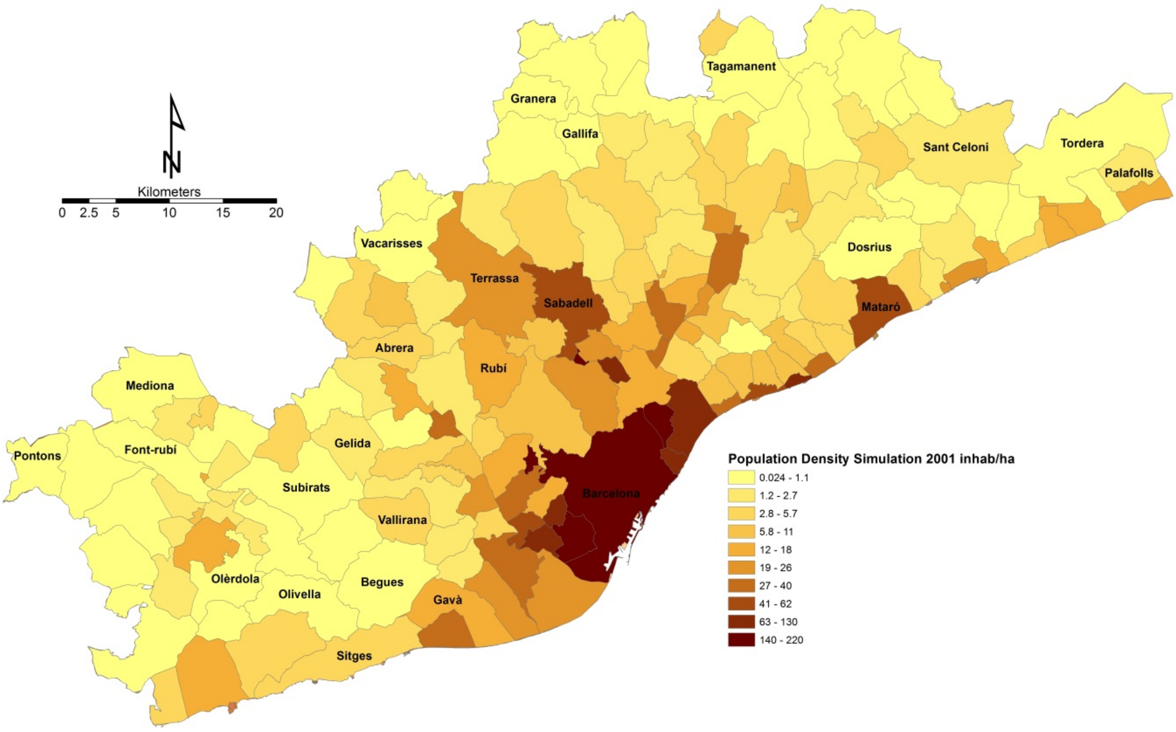


Figure 6.10 Simulation map of population distribution in MAB in 2001, inhabitants per hectare

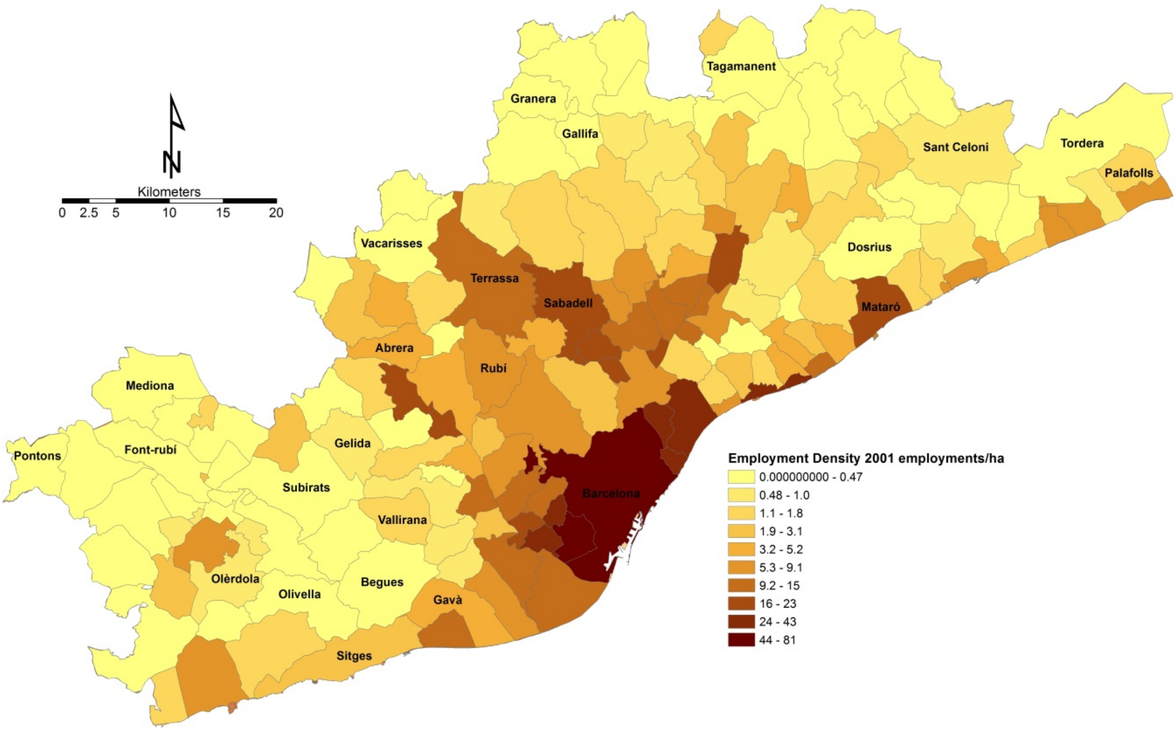


Figure 6.11 Reference map of employment distribution in MAB in 2001, inhabitants per hectare

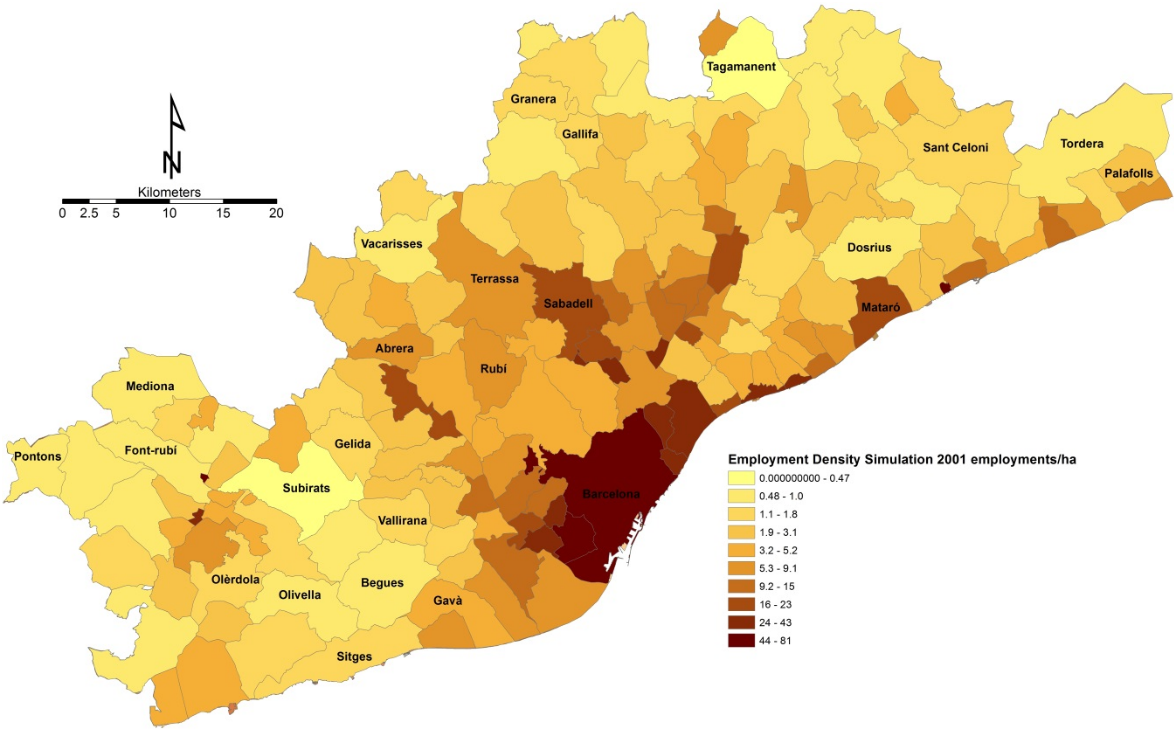


Figure 6.12 Simulation map of employment distribution in MAB in 2001, inhabitants per hectare

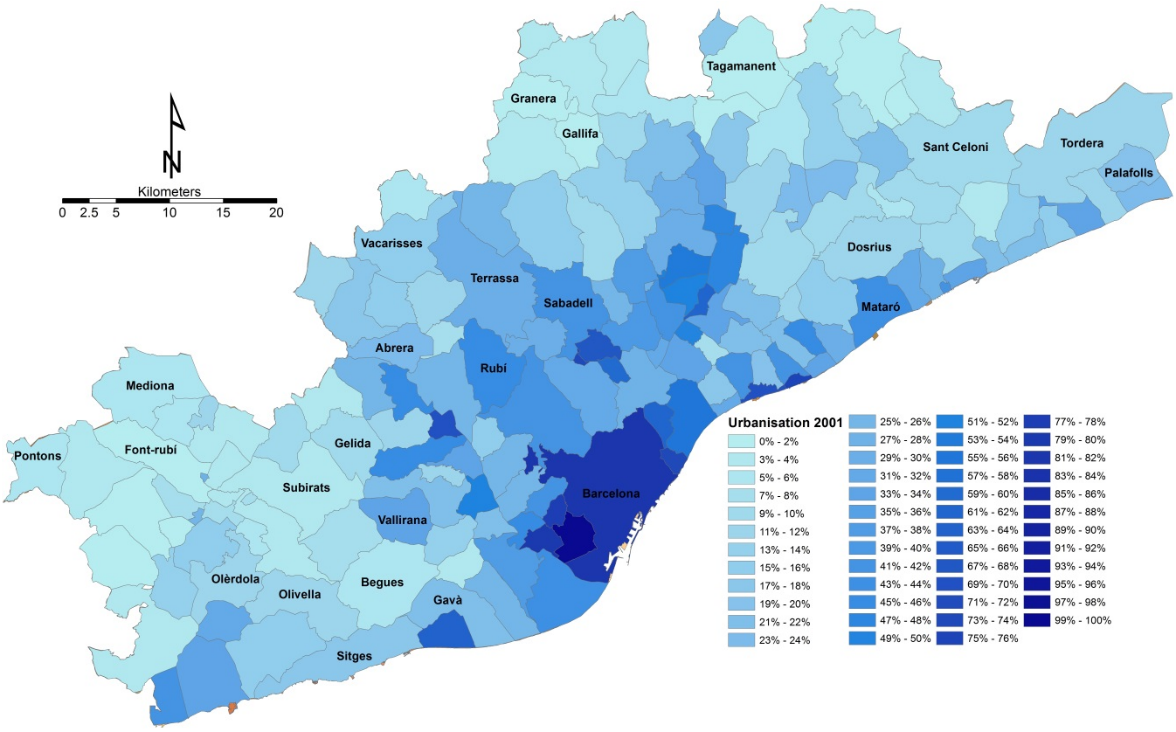


Figure 6.13 Reference map representing cell states (aggregated urbanised land) for 2001

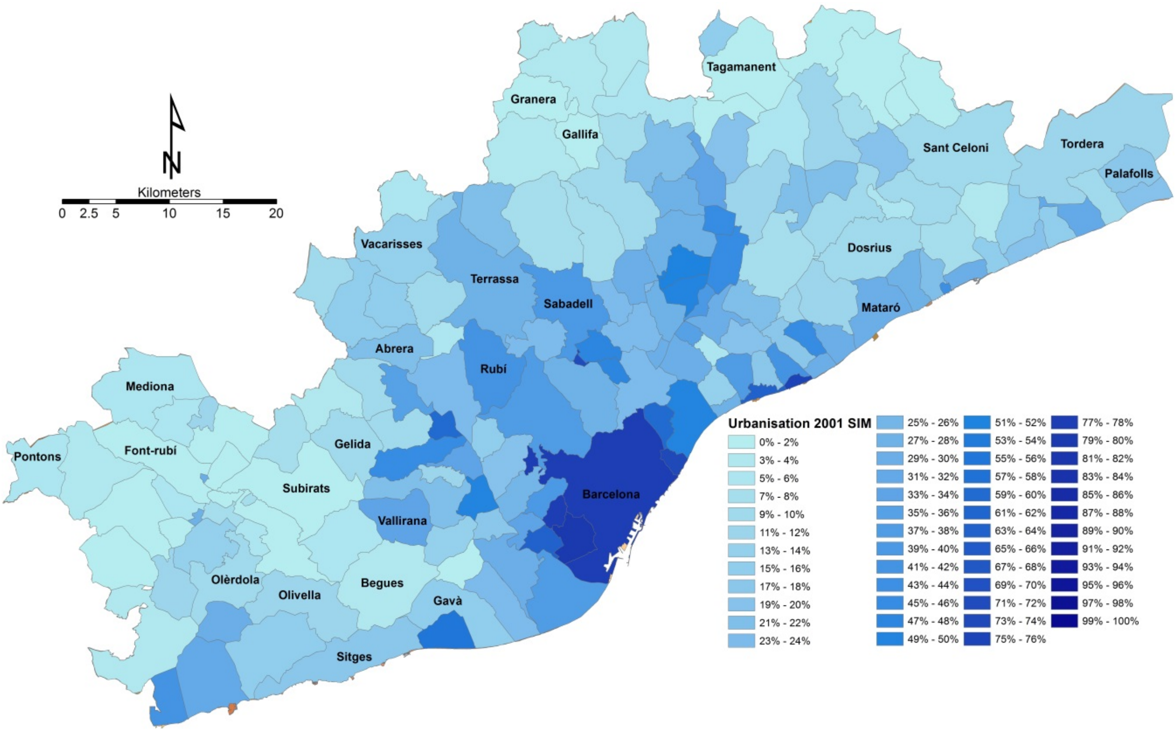


Figure 6.14 Simulation map representing cell states (aggregated urbanised land) for 2001

6.2.2 Model application

The model was able to produce a fairly competent simulation that is expressed by the relatively high value of the performance measure, the value of *kappa* of 0.427 which represents a moderate agreement if compared with other applications of micro-scale models reported in previous chapters. This value for agreement corresponds to a much higher value of 0,991 if area (and not cell states) was to be used, showing that the difference between cell states in both simulation and reality is considerably small, despite the low *kappa* value. Figure 6.15 shows this difference and it is possible to see that the larger values of difference occurred in areas of large urban land demand around the city of Barcelona and in some of the second tier municipalities in the central and coastal areas of MAB.

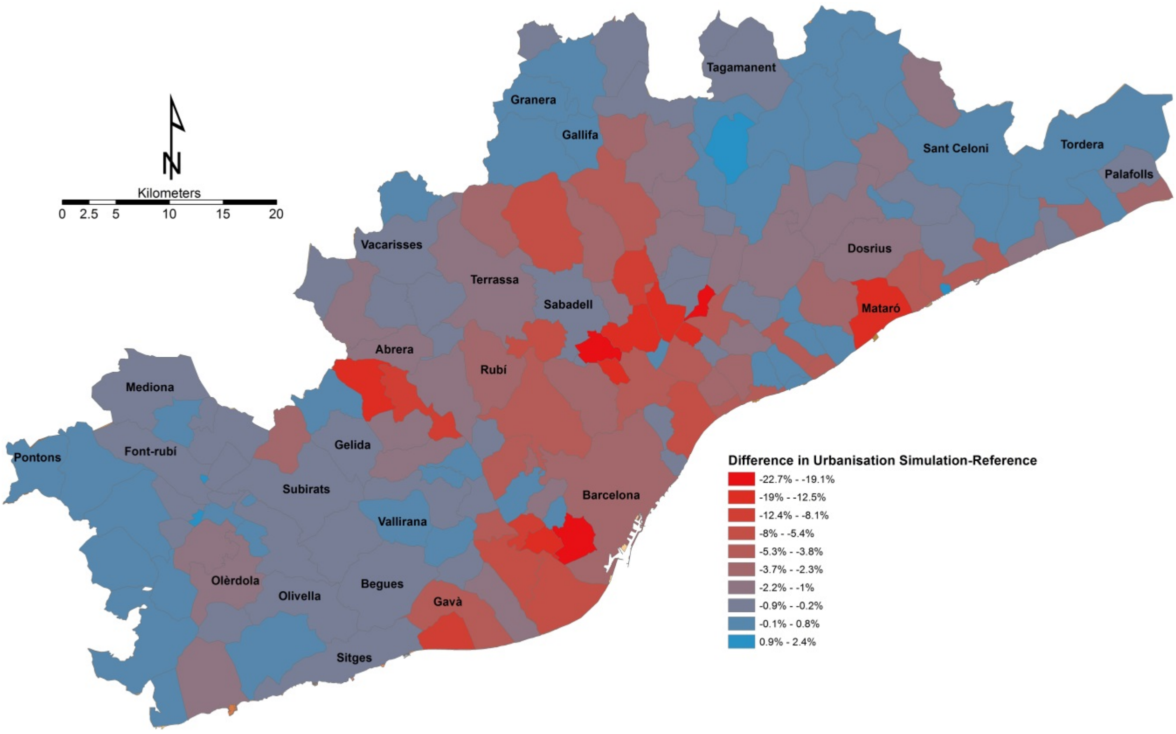


Figure 6.15 Difference in urbanisation (cell states) between simulation and reference maps

However, this level of agreement can be considered encouraging as it still is higher than common values of agreement for traditional CA models that focus on large regions such as the MAB.

6.3 Conclusions

The macro-scale model aims to simulate the evolution of land use demand by modelling the areas of urbanized land at the municipality level as a function of the location of population and employment, considering accessibility measured using the road network to effectively integrate the land use and transport interactions at the regional/metropolitan scale. The macro-scale model gives indications that it can be used to model these scale of spatial interactions. However, the relative low level of accuracy of the model despite a fairly competent overall performance, as indicated by the level of agreement between simulated and reference maps, indicates that some issues must be the focus of further research. The transition potential measured by a gravitational function was used to give the model some level of comparability with other regional spatial interactions models usually used coupled with traditional CA models, as it is the case of MOLAND (Engelen et al., 2002). Further research will be done to compare in a more accurate way the performance of this CA model with those other types of regional interactions models using compatible datasets. The consideration of other modes of metropolitan transport, namely the train, will also improve representativeness of the model in terms of how accessibility is calculated. This comparison is only constrained by the existence of feasible datasets for every component of the model. The consideration of a logit function to create a balanced distribution function of both population and employment has proven to be correct if compared with other implementations of CA, as the one presented in chapter 4.

The main aim of designing simple CA models that can be applied to different data-rich contexts led to the consideration of the simplest datasets that could be a common minimum denominator to those other contexts.

This model can be used to illustrate the possibilities that CA concepts have to simulate aggregate consumption of land at a municipal scale as well as future population and employment distributions. The overall values of urbanized land per municipality are meant to be considered as land use demand at the micro-scale, and will be used as a constraint to a more traditional, local scale CA model such as the one presented in chapter 5. Other factors such as quality of local environment, local policies for attracting residents and jobs, or multimodal accessibility systems are among the future lines of research in future developments of the model.

7 A Multi-Scale Cellular Automata Model

This chapter presents the multi-scale CA model that aims to simulate the dynamics of land use change in different scales by capturing relevant interactions that occur and can be modelled and parametrised at different scales. The model makes use of the two models previously presented in chapter 5, the micro-scale CA model, and in chapter 6, the macro-scale CA model. The literature shows no records of this type of integrated multi-scale approach using CA models. As discussed in chapters 5 and 6, there is a recent discussion about the issue of scale, mainly with high resolution models at the local scale (e.g. Stevens et al. (2007)), and variable cell sizes that can include the entire study area as the work of White et al. (2015). This highlights the novelty of testing the integration of these two levels of modelling resolution using the CA concept at both scales.

Section 7.1 presents the structure of the model, drawing on the mathematical and conceptual formulations presented in the previous chapters. Section 7.2 describes the application of the model to the case study of the Metropolitan Area of Barcelona (MAB), where the entire MAB is taken as the case for the macro-scale model whereas some municipalities are selected as cases for the application of the micro-scale model. Section 7.3 presents the main results that include the calibration of the models for each case study and a discussion about the validity of the multi-scale model. Finally, section 7.4 presents some conclusions about this type of modelling and its potential and limitations.

7.1 The multi-scale model

The multi-scale CA model is an integration of the two other CA models developed and presented previously in this dissertation. The multi-scale CA model aims to simulate land use dynamics considering how these dynamics operate at the two relevant scales of analysis, the metropolitan/regional analysis and the local scale land use dynamics.

The model uses the calibration that both micro- and macro-scale CA models produce to capture those dynamics. It then integrates the two models into one single system that aims to simulate the decisions about overall land use change at a municipal level (the macro-scale) and how this land use change will occur in a more traditional land use allocation model at each municipality (the micro-scale).

This multi-scale CA model follows the same principle of generalisation as the two contributing CA models. These models aim to be applicable to all local or planning contexts by using a pragmatic approach in terms of data requirements and representativeness. Therefore, these models use available spatial units and associated datasets that can be found in different locations without the need for extra pre-processing of data or further customisation.

This is recognised and assumed as an important design option, and it implies a certain degree of loss of representativeness of the results to ensure a uniform methodological CA proposal.

Figure 7.1 represents the conceptual architecture of the multi-scale CA model.

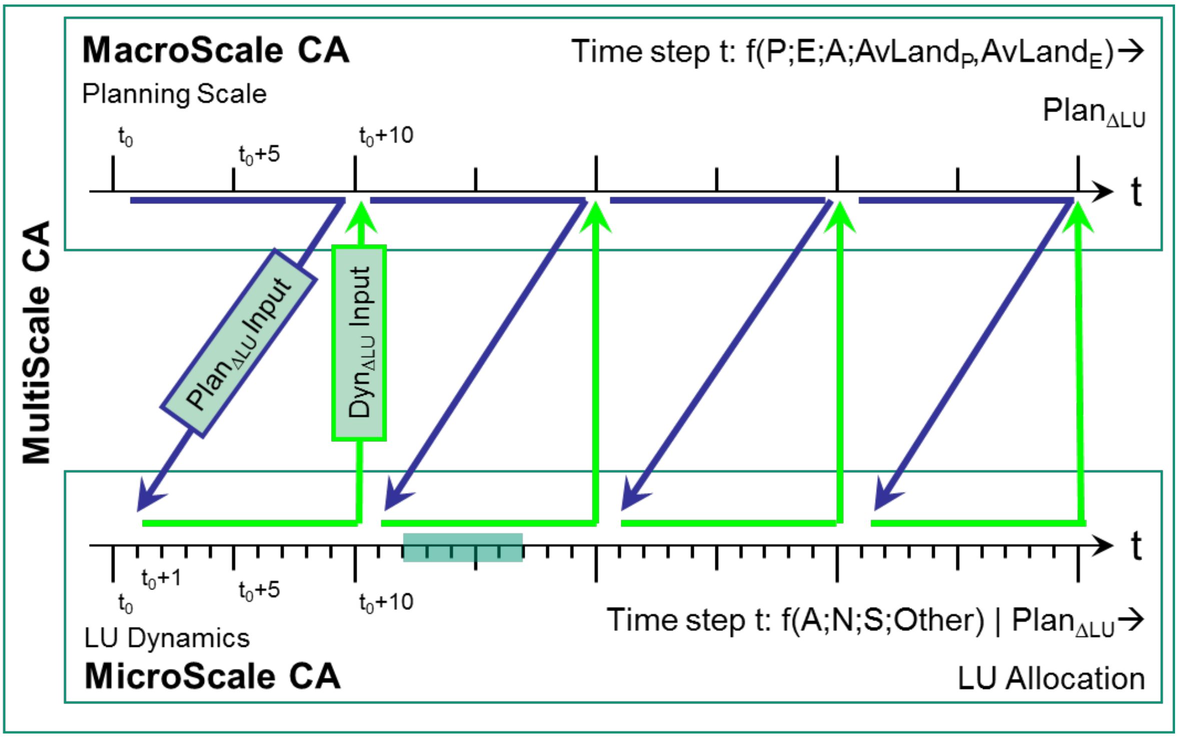


Figure 7.1 Conceptual architecture of the multi-scale CA model

The macro-scale component aims to simulate the planning scale, where major decisions on land consumption are made considering a more strategic level of decision making, taking into account important dynamics that can be observed in a given territory at a given moment. Population and employment distributions can be used as the major drivers at this scale. Accessibility also plays a role as a driver that takes into account the previous two variables and the way the higher level macro-scale infrastructure services location choices.

Due to the competitive (yet combined) planning process associated with a metropolitan area, spatial interaction of these last variables as competitive parameters in regional competitiveness define a neighbourhood interaction that justifies the application of the CA concept.

The micro-scale component aims to simulate land use change within municipalities. Once planning decisions are made for the municipality in the context of a regional/metropolitan planning process, hence generating aggregated values of land consumption, a process of land use allocation starts in each municipality, usually in the form of regulatory zoning maps or equivalent planning tools. These local processes take place considering some constraints as land suitability or planning regulations. The main explanatory variables at this scale are again population and employment distributions, combined with local accessibility. These variables are used at a very fine grain, corresponding to a cell structure that represents better local urban form. However, other important factors and land use interactions play a major role at this scale, justifying the use of CA as the main modelling concept.

Both the macro- and the micro-scale CA models were designed to use existing spatial units that combine reliable information (from censuses and other official sources) with regional administrative (macro-scale) and urban (micro-scale) forms. The macro-scale model uses municipalities as cells. Municipalities are the main decision making spatial unit in planning, deriving their land use policy from their existing masterplans. The micro-scale model uses census tracts to model land use dynamics; the use of census tracts allows the model to combine urban form with the data that underlines change.

Cell states are aggregated classifications of land use at both scales. The macro-scale CA model tries to simulate the overall value of artificialized land for each municipality, an important planning parameter in the definition of a planning strategy in the context of competitive (and yet collaborative) metropolitan areas. The micro-scale CA model uses a set of six aggregate land uses that model the main urban uses (taking into consideration density) as well as the industrial ones. This classification identifies the available empty land for development for both urban and industrial uses, and considers a final land use that combines all the land that cannot be urbanised by many different reasons, such as agriculture capacities, ecological value, geomorphological or topographical constraints, just to name a few.

Neighbourhood distance is a calibration parameter for both the micro- and the macro-scale CA models. Due to the different nature of the spatial interactions, neighbourhood has a dual formulation at the macro-scale where there is an outer distance that bounds the spatial interaction between cells, along with an inner neighbourhood that simulates positive or negative spill overs of the potential of a given cell to its neighbours.

Transition rules have different formulations. At the macro-scale a gravitational based function is used to combine cumulative effects of population and employment distributions with accessibility levels, being all of these parameters calibrated to control their weight on the final function. At the micro-scale, a composite value of transition potential that includes accessibility, land suitability and neighbourhood interaction between land uses is used to establish the potential of change at a given moment. Both scales use a logit function to normalise the respective values of transition, so that the poll for selecting cells that change at each time step is not biased by cell size.

Finally, time resolution also varies in both models. The macro-scale CA model operates over periods of 10 years, simulating the time horizon for planning in the majority of planning contexts. The micro-scale CA model provides options for simulating time sets of 1, 5 or 10 years, to simulate local conditions of urban development.

7.2 Application to the Metropolitan Area of Barcelona

The multi-scale CA model was applied to the case of the Metropolitan Area of Barcelona (MAB). This case study was selected due to its natural complexity and diversity of urban dynamics. MAB has experienced a strong urban growth dynamics in the recent past due to the increase of attractiveness that Catalonia (with a striving economy, especially in the pre-2007 crisis period) and MAB (as its major urban spearhead) has exerted at a global scale, both as a primary destination for immigration influxes in the 2000s and as an important European regional metropolitan area.

The macro-scale CA model was applied to the entire MAB, modelling the land use dynamics using population and employment distributions considering metropolitan accessibility (with just the car mode being considered).

The micro-scale CA model was applied to three selected municipalities in MAB to simulate land use dynamics at a fine grain resolution, considering a constrained land use demand (obtained from the macro-scale CA model) and population and employment dynamics at the local scale of each municipality. The selection of this limited number of cases as instances for the micro-scale model followed a series of criteria that were design considering model limitations and representativeness of the modelling exercise.

The first constraint to this application is the computational limitation of the software implementation of the model that due to its architecture cannot cope, at this stage, with a full application to the whole of the 164 municipalities of MAB. These computational limitations arise from the way the software was designed and the way it manages memory allocation, limiting its ability to scale up.

Nevertheless, the CA formulation and the multi-scale mechanisms are fully implemented in the model. The computational limitation lies (for now) in a limited number of overall variables (and their overall memory allocation) that operate in the software. This is a limitation that can only be tackled with the support of an expert on software design to optimise these important features, which is an immediate future line of research.

Therefore, micro-scale case studies need to show different land use dynamics due to their different importance in the polycentric system of MAB. These should also show different sizes, land use availability, different locations in MAB that could also represent different natural features.

The selected cases are the municipalities of Barcelona, for its importance and role as the spearhead of MAB, Mataró, a coastal municipality that illustrates the expansion of MAB core urban area towards the Northeast along the attractive area of the coast line, illustrating a current pressure in Catalonian and Spanish urbanisation processes, and Terrassa as a dynamic large municipality in the inner periphery of MAB playing a role in the competitive dynamics of the metropolitan area.

7.2.1 Macroscale case study – the Metropolitan area of Barcelona

The Metropolitan Area of Barcelona (MAB) was already introduced in chapter 6 but it is useful to provide a new quick introduction to put the case study into context. MAB, the large city region spearheaded by the city of Barcelona, is the metropolitan capital of Catalonia, one of the autonomic regions of Spain. MAB is composed by 164 municipalities which vary considerably in area, population, and employment. Municipalities in Catalonia hold legal powers for spatial planning but have different technical and financial resources to develop and implement their plans and other local policies due to the great variation in size and population. Figure 6.2 depicts MAB and shows how municipalities vary in size, with population varying from few hundreds of inhabitants up to a million and a half, the same happening with employment. Figure 6.2 also locates the three selected municipalities that are modelled by the micro-scale model. A longer description of the MAB can be found in chapter 6, section 6.2.1.

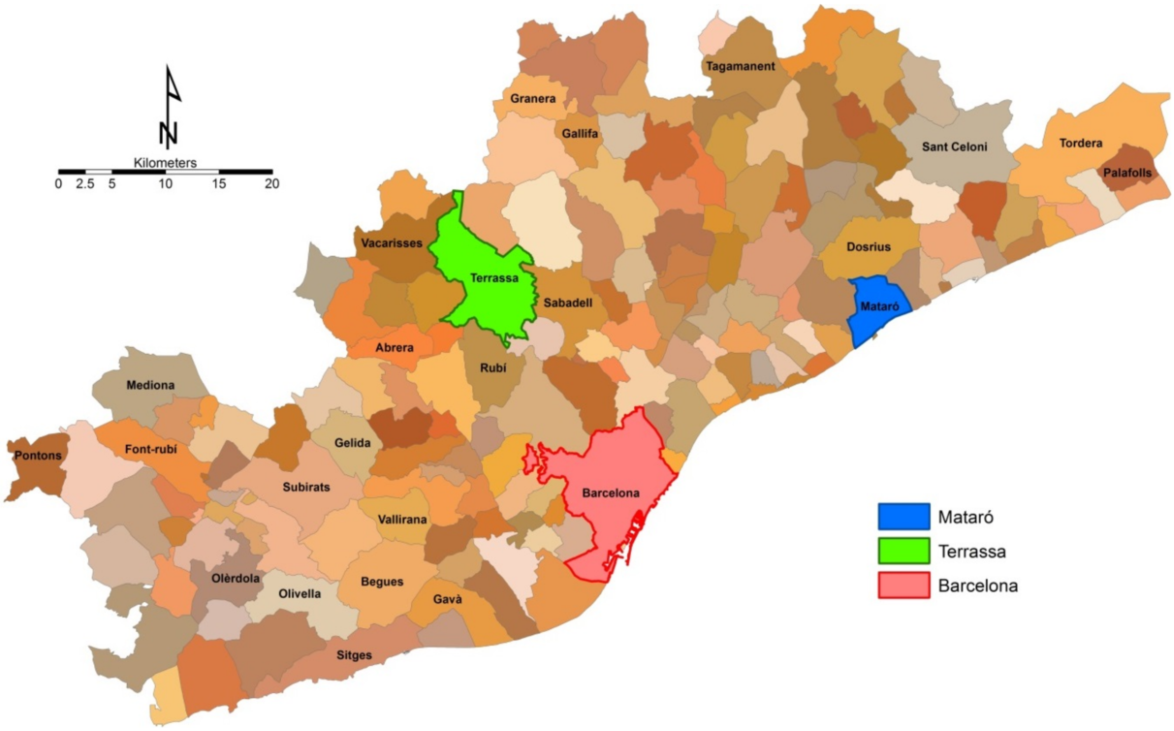


Figure 7.2 Metropolitan Area of Barcelona (MAB)

7.2.2 Microscale case studies

Barcelona is the largest city in MAB, with approximately 1.7 million inhabitants in 2001. It is limited by growing neighbouring municipalities of Hospitalet del Llobregat to the Southwest and Badalona to the Northwest, being naturally constrained by the natural park of Collserola to the north and the Mediterranean sea to the south. Barcelona has experienced a slower growth dynamic in the recent past due to the reduced available area for urban expansion (a more common phenomenon in the last decades)

Mataró is a smaller municipality of around 100 thousand inhabitants (in 2001) located in the coastal area that plays a dual role within the polycentric structure of MAB. It is one of the municipalities that conjugate the peripheral character, being more attractive to residents due to lower prices than the core of MAB, while offering the amenities of the coastal area, with an increased attractiveness in comparison with Terrassa.

Terrassa is one of the most important municipalities in terms of urban functions in MAB with around 160 thousand inhabitants (in 2001) located north to the city of Barcelona, after the first natural barrier of the Collserola Park. This is one of the most important peripheral cities in MAB, being an attraction pole for the land use demand that cannot locate in the city of Barcelona or under higher land costs as in the more coastal municipalities.

Mataró and Terrassa present similar levels of accessibility, although the latter is more central not only in geographical terms but also as an employment centre located within one of the most important cores of industrial (hence employment) location.

7.3 Discussion of results

The micro-scale CA model was used to calibrate the case studies of Barcelona, Terrassa and Mataró in order to obtain the best set of parameters that will be the input for the multi-scale model. Conversely, the macro-scale CA model was also used to calibrate the case study of MAB as the macro-scale level in the multi-scale model. All three case studies used datasets for 1991 and 2001, which coincide with the dataset used by the macro-scale CA model. The performance indicator is the same one described in chapter 4 and used in for both the theoretical instances (chapter 4) and the Coimbra case study (chapter 5).

The case of Barcelona was able to achieve a value for *kMod* of 0,746. The correspondent reference and simulation maps for land use are depicted in Figure 7.3.

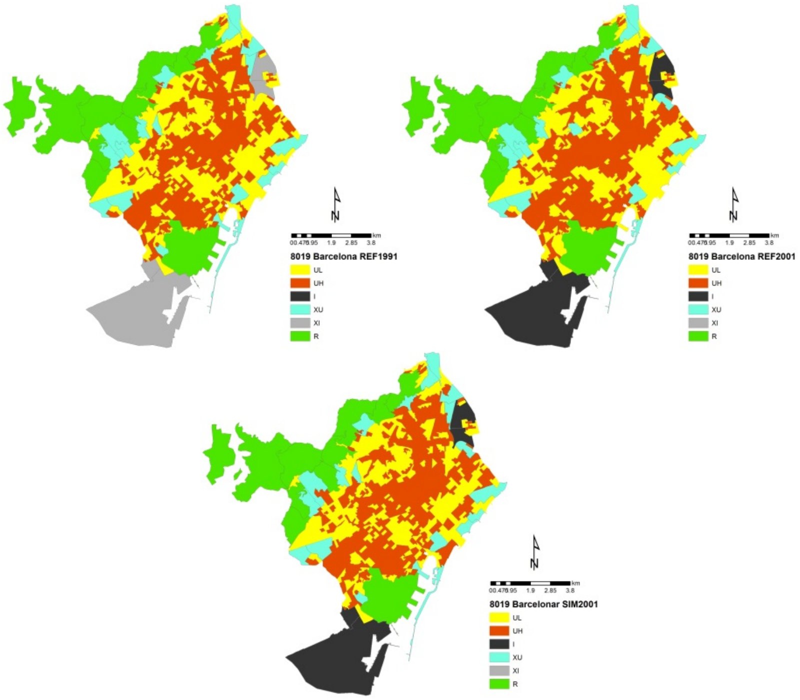


Figure 7.3 Calibration land use maps for Barcelona: (a) reference map 1991, (b) reference map for 2001, (c) simulation map for 2001

The case of Mataró achieved a lower value for *kMod* of 0,632, yet still considered as a very good level of simulation. The correspondent reference and simulation maps for land use are depicted in Figure 7.4.

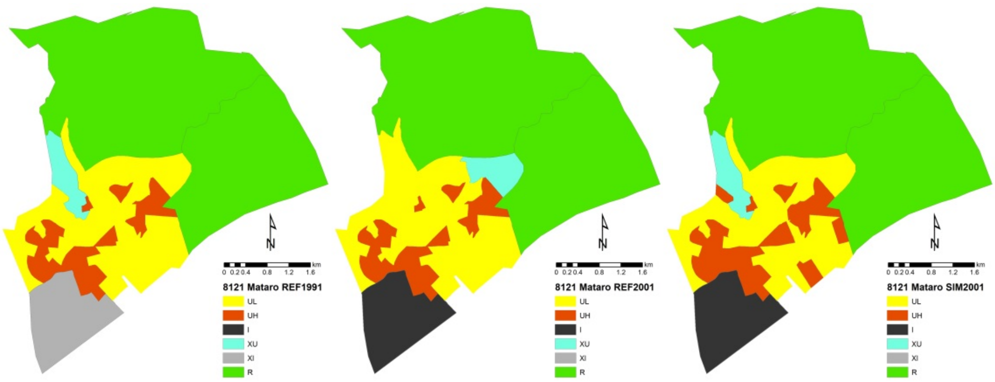


Figure 7.4 Calibration land use maps for Mataró: (a) reference map 1991, (b) reference map for 2001, (c) simulation map for 2001

Finally, the case of Terrassa was able to achieve also a significant a value for *kMod* of 0,687. The correspondent reference and simulation maps for land use are depicted in Figure 7.5.

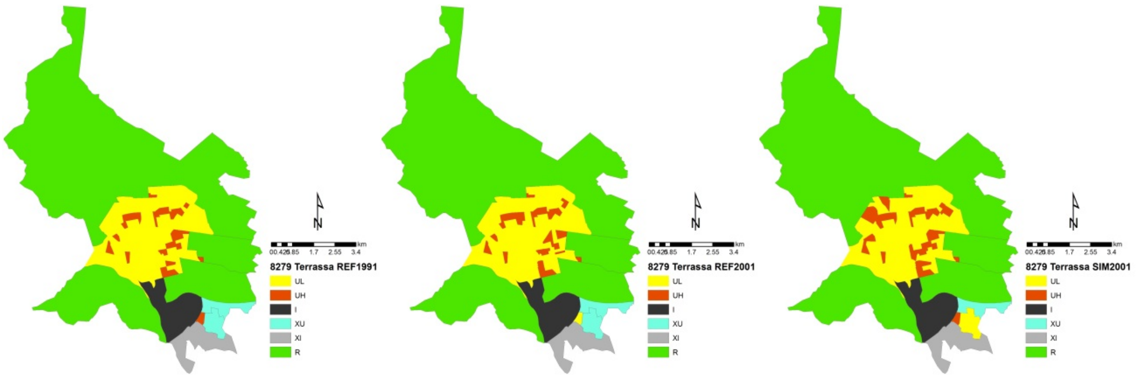


Figure 7.5 Calibration land use maps for Terrassa: (a) reference map 1991, (b) reference map for 2001, (c) simulation map for 2001

All the three runs of the micro-scale CA model have performed in line with all the other previous applications of the model, consolidating the ability of this CA implementation to model small urban areas.

The calibration of the macro-scale CA model for the MAB and the discussion of its results are available in section 6.2.2 of the previous chapter. The results were not as good as the ones obtained for the micro-scale CA model in terms of the absolute value for the performance indicator but are still of the same order as the ones available for comparable models that operate at the same scale.

To test the capabilities of the model a simple scenario was designed to simulate the growth that might occur on MAB and on the three selected municipalities during the period 2001 to 2011, keeping similar growth rates for the period 2011 to 2021. Due to the experimental nature of the macro-scale CA model there is no expectation that the results should achieve a high degree of reproduction of the reality in 2011, as the quality of the simulation for both the micro- and macro-scale CA models still does not provide a full validation of the model against the observed reality in this very dynamic and complex urban system.

The scenario is defined by a population growth rate that mimics the overall increase of approximately 387 thousand inhabitants in MAB during the period of 2001 to 2011, which corresponds to an overall increase of 8.0%. The decade after is considered to reflect the effects of the crisis and a lower growth rate of 2.5% is considered. The same behaviour is taken into account for employment, with a steep 12% increase rate corresponding to the employment boom in MAB during the 2000s, with a strong deceleration for 4% in the next decade. Accessibility conditions, measured via travel times in the uncongested road network, are considered constant due to the reduced importance of road investment in the most important links of the network.

Table 7.1 Macro indicators for population and employment growth and accessibility conditions

Variable	2001-2011	2011-2021
Population	8%	2.5%
Employment	12%	4%
Accessibility (infrastructure)	Same conditions across MAB	

Figure 7.6 and Figure 7.7 represent, respectively, the distribution of population and employment for 2011 across MAB (in inhabitants per hectare and employments per hectare, respectively), while Figure 7.10 and Figure 7.11 depict the same variables in 2021. Figure 7.8 represents the distribution of the urbanised land in MAB in 2011 (in percentage of the municipal area) and reads along with Figure 7.9, where land use maps are depicted for the three municipalities for 2011, showing local land use allocation considering the overall increase of land for each municipality. The same happens with Figure 7.12, which depicts the distribution of the urbanised land in MAB in 2021 in combination with Figure 7.13, which again presents the land use maps for the three local case studies.

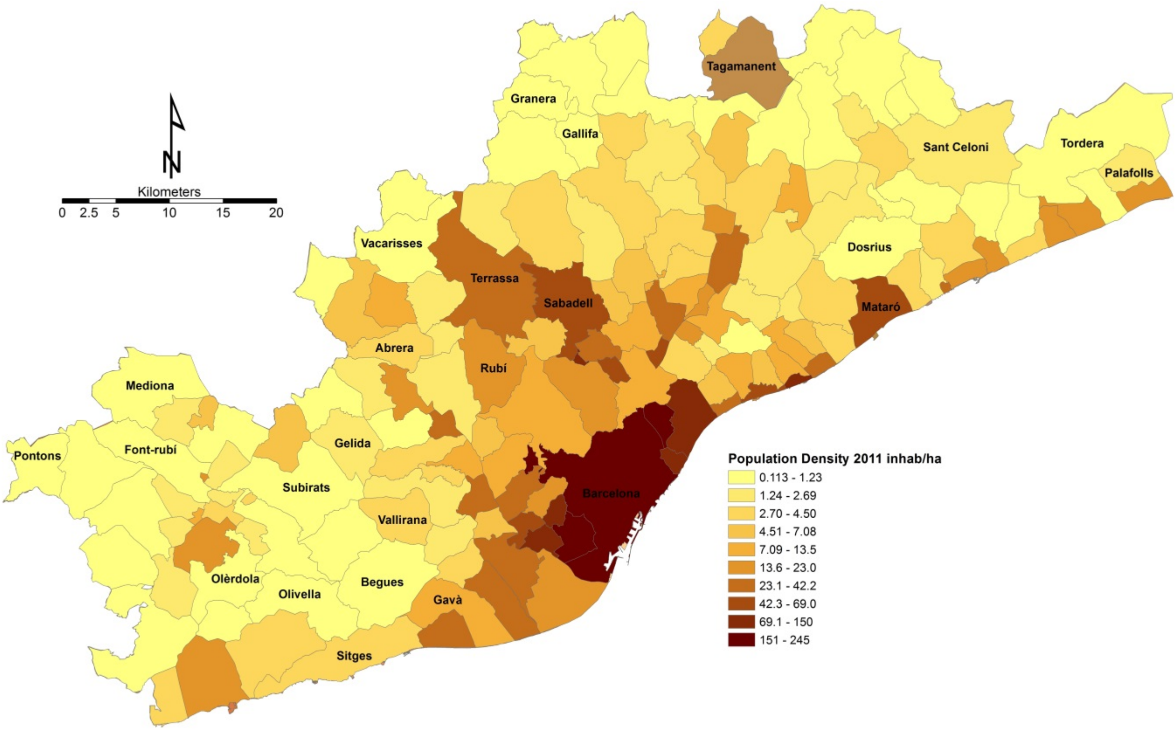


Figure 7.6 Population distribution in MAB for 2011, inhabitants per hectare

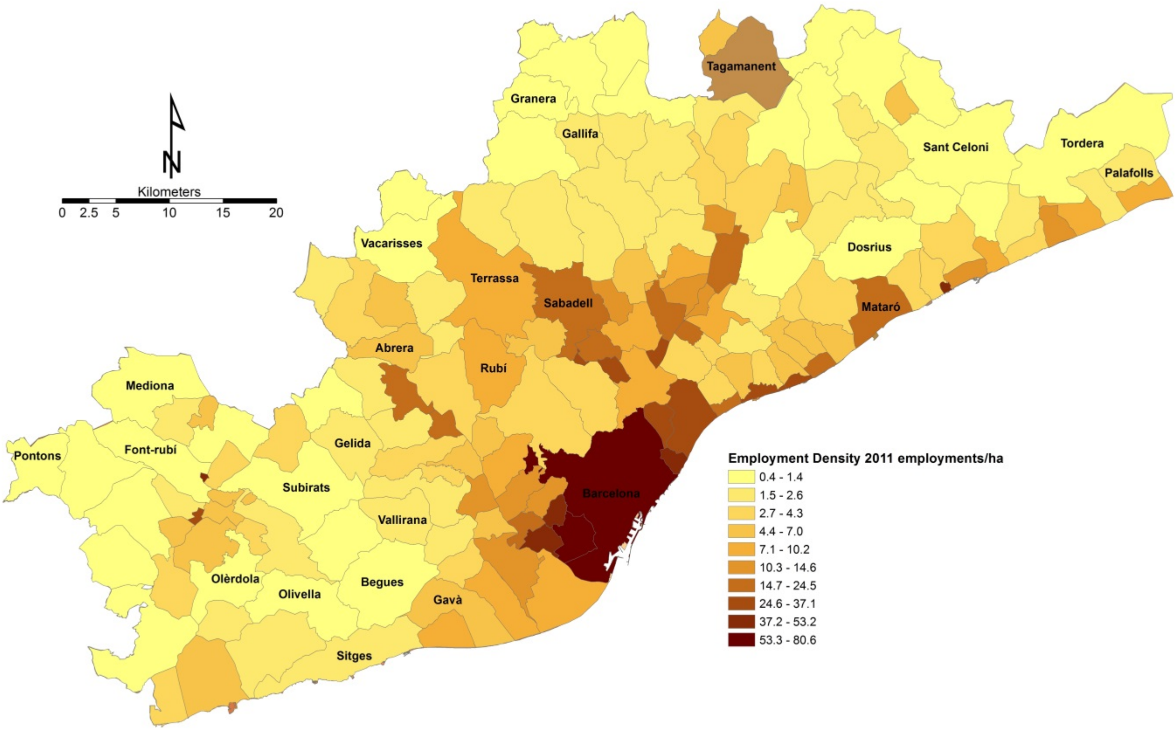


Figure 7.7 Employment distribution in MAB for 2011, employments per hectare

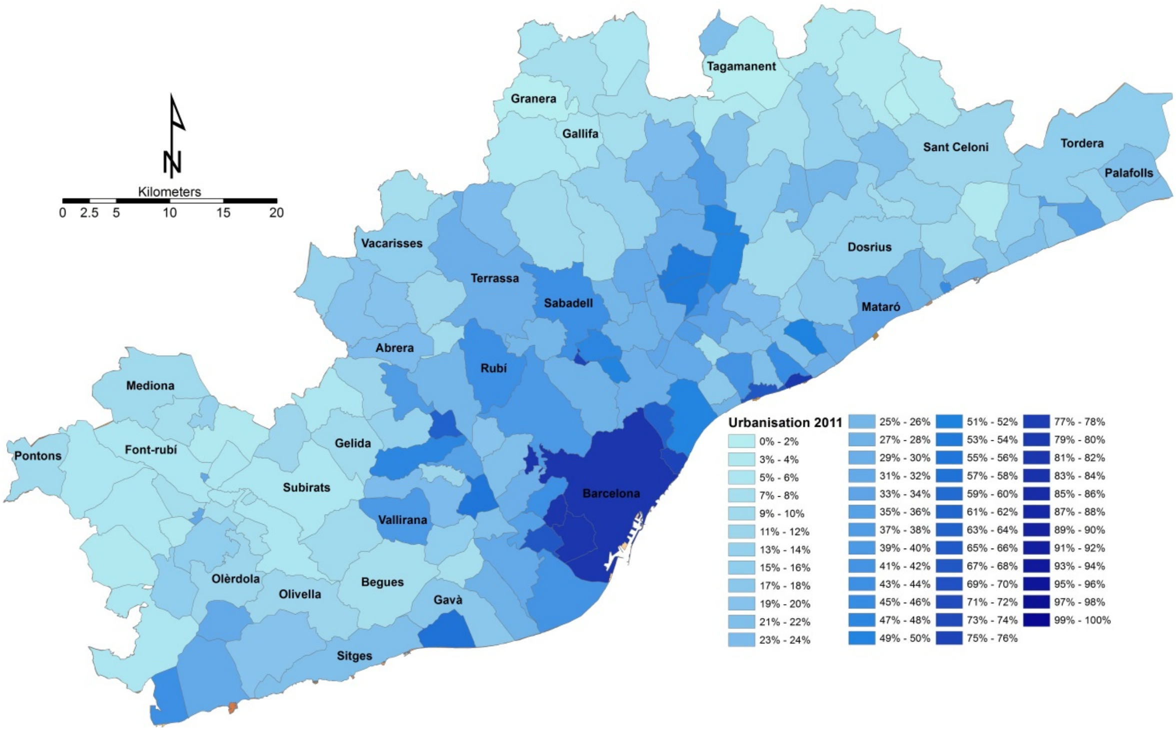


Figure 7.8 Artificial land distribution in MAB for 2011, percentage of total municipal area

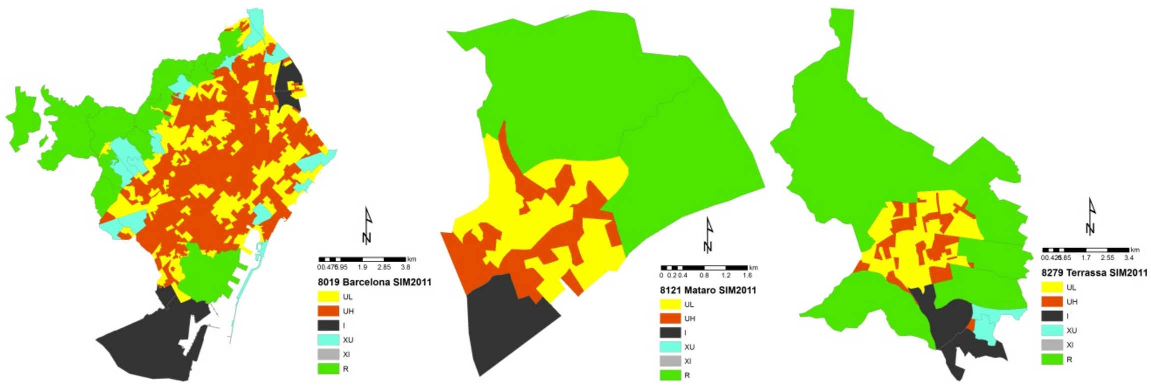


Figure 7.9 Land use maps for (a) Barcelona, (b) Mataró and (c) Terrassa for 2011

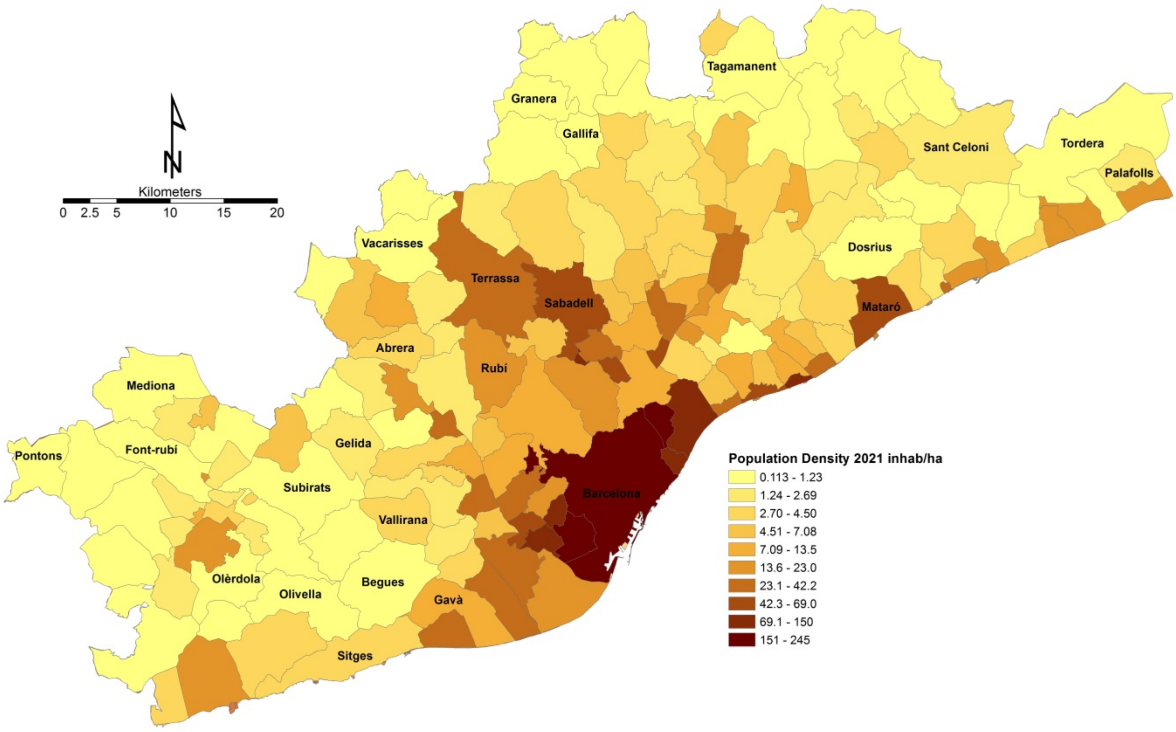


Figure 7.10 Population distribution in MAB for 2021, inhabitants per hectare

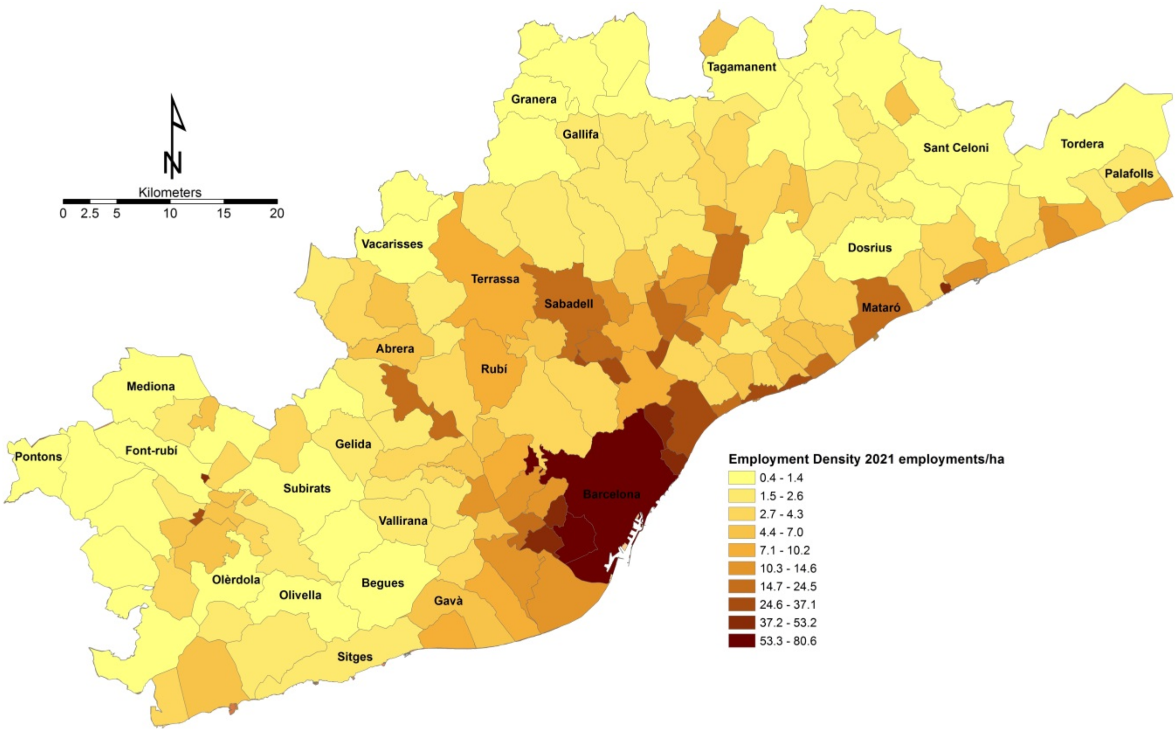


Figure 7.11 Employment distribution in MAB for 2021, employments per hectare

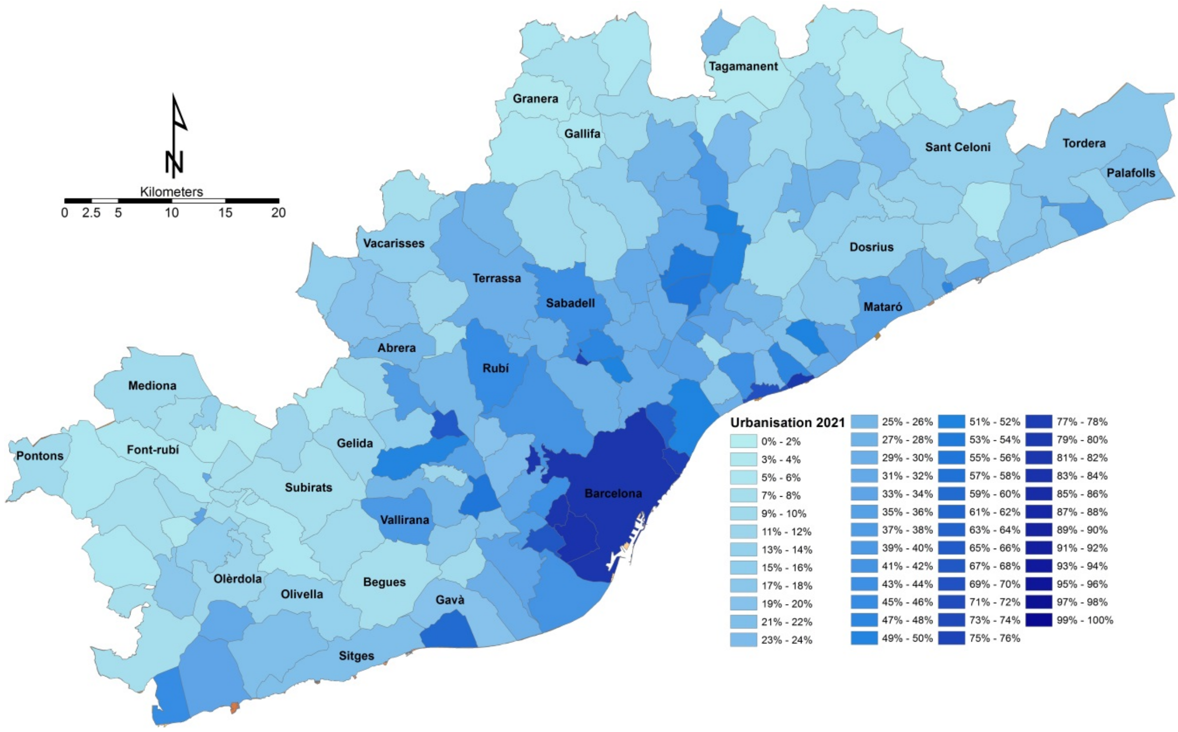


Figure 7.12 Artificial land distribution in MAB for 2021, percentage of total municipal area

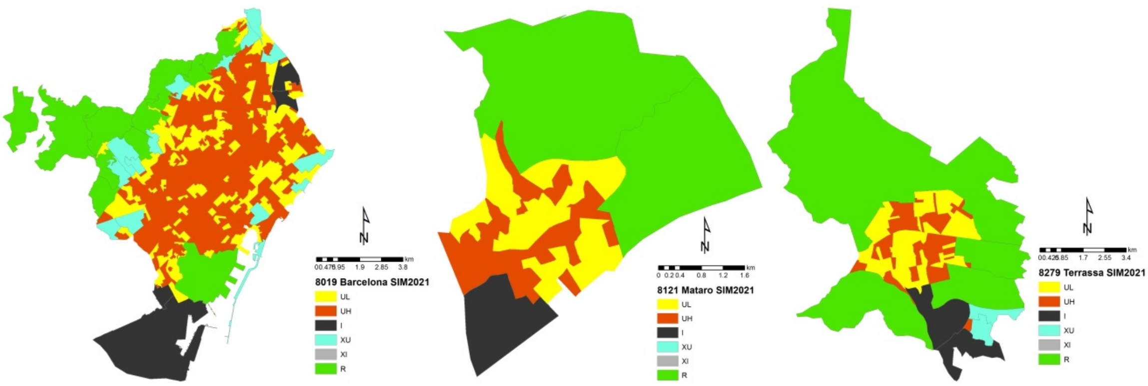


Figure 7.13 Land use maps for (a) Barcelona, (b) Mataró and (c) Terrassa for 2021

The model was able to deliver an even distribution of the increase of population in the periphery of the metropolitan area, possibly favouring the impact of good accessibility conditions and replicating the lack of capacity that the central municipalities of Barcelona and close neighbours might evidence in the near future due to lack of available land for development. The average increase in population drops from 10% in the first decade to 4.1% in the second (with maximum values decreasing less).

The evolution of employment also slows down but with a slightly different distribution. The average growth rate drops from 6.1% to 2.8%, but with maximums of 34% dropping to an as much impressive 25%, illustrating the existence of preferred locations for job concentration.

This distribution of population and employment is followed by an expected even increase of newly artificialized area across MAB (considering the dependency of these variables), with an average of 2.1% increase (maximum of 4%) in the first decade and a slower 1.1% increase (and maximum of 2%) in the second one. This may indicate a balance between the main drivers of land use change, as no municipality really excels against the rest.

At the micro-scale level, the three municipalities modelled experience similar trends with different intensities.

Barcelona experiences an increase of 1.1% and 1.5% of urbanised area in the two decades, for an increase of population of 7% in the first decade dropping to just 2% in the second. Employment suffers a stronger deceleration from 14.7% to just 2.5% in the second decade. All these values are lower than the correspondent values for MAB (except of employment growth in 2001-2011), a possible sign of the saturation that the central and most occupied municipality of the metropolitan experiences.

Mataró seems to have a different capacity to attract people and employment, experiencing more urbanisation. The population of the coastal municipality increases 10% in the first decade slowing down to 3% in the second, while experiencing a stronger growth in jobs, with 21.4% in the first decade and a still high 14.9% in the second period. This strong demand corresponds to more 3.6% of urbanised land in the first decade against just 1.6% increase in the second.

Finally, Terrassa is somewhat in the mid-term between the two previous municipalities. It experiences 10% of population growth in the first decade reducing to 4% in the following one (in line with Mataró, also a peripheral municipality), while employment grows slower, with 16.7% and 3.3% respectively. The correspondent increase of urbanised area is of 1.6% and 1.3% for the two periods respectively.

7.4 Conclusions

These results show a feasible exercise of simulation of land use change considering accessibility and considering simple (but not simplistic) scenarios of future evolution. The results are quite interesting as they contribute to give evidence that the CA modelling concept is fit to simulate these phenomena in a multi-scale context, without any loss of conceptual character.

The multi-scale CA model presented is an integration of two different CA models that operate at the local and regional/metropolitan scales. This integration was done considering the prospective component of the models, as there are current computational constraints that made the full implementation at the calibration stage unfeasible. This implementation could not operate over the entire territory under study, mainly because of memory management and limits to embedded loops and cycles that are still constraints of the Visual Basic 6 (VB6) technology, and that are beyond the candidate's current programming skills. This lack of memory availability for very large memory needs for both the CA modules and for the brute force computing calibration procedure based on particle swarm (PS) hampered the possibility of a full calibration of the two scales. At the same time, the experiments done for testing smaller datasets showed that a strong divergence occurred from simulated outputs to the observed reality, leading to a very poor level of calibration. This is the main issue of future research on the multi-scale CA, as there is a potential of developing a step-wise, fine grained temporal process of calibrating the models, provided that a proper software architecture is designed making use of a different, fully object-oriented programming language with no capacity limits.

Nonetheless, the current multi-scale CA model is valid and responds to the aim of this research to model land use change at multiple scales. The current implementation makes use of the calibrated set of parameters obtained by both the micro- and macro-scale CA models, capturing spatial interaction and operating different transitions rules that are design specifically for each scale, with different aims. Accessibility is considered at both scales with different influences on the transition rules, cells are designed to capture form at each scale, neighbourhood represents different types of interactions and transition rules are defined considering the drivers at each scale. There are still relevant capacity limits deriving from the software implementation in VB6, hence the use of only three municipalities as local scale case studies, as the software cannot cope with a full run with all municipalities. The concept is nonetheless proven and the results are encouraging. The multi-scale CA model inherits results from the full CA formulation from both the micro- and the macro-scale models, complying with the goal of using robust CA formulations in all models. The multi-scale model provided feasible future outcomes that depend mainly on the future scenarios that are designed by the user.

8 Conclusion

This dissertation reported the research done on cellular automata (CA) modelling for land use change, considering the ongoing topical subjects of scale, cell form, neighbourhood interaction and transition rules. A theoretical framework for urban modelling and urban studies and for the use of CA models in urban modelling was discussed and used to support the identification of the most relevant subjects of research. A series of models were developed accordingly and their main features and results presented, illustrating both their potential and their limitations.

This chapter is dedicated to the final critical discussion about the main findings, issues and potential that outcomes from this research. Section 8.1 presents a series of conclusions about the many research outcomes, focusing on the details of the models. Section 8.2 is dedicated to a reflection on the research process and on its outcomes. Finally, section 8.3 presents the main avenues of future research that the candidate will pursue in the short-term and in the mid- to long-term futures, consolidating a research agenda on multiple applications of CA in the broad remit of urban studies.

8.1 Moving cellular automata modelling forward

The research proposed to develop a series of components of CA models that are not fully explored in the existent literature.

The inclusion of accessibility as an endogenous component of the CA models (common to all scales) is a valuable achievement of the research. The definition of a simple accessibility model based on the road transport network considering its capacity levels is a step forward as there are little (if any) examples of this integration in the literature. The accessibility model can be easily expanded for a multimodal approach, depending only on the computational implementation. The fact that accessibility parameters are calibrated along with the land use parameters brings a new level of understanding of the effective land use and transport interactions using CA models, demonstrating the potential of this concept as an effective tool to deal with what can be considered one of the most studied problems in urban and transport studies. This achievement is expected to be even more useful as a decision support tool in the context of less capacitated planning bodies (e.g. municipalities) in less capacitated or regulated planning contexts (e.g. developing countries).

The use of irregular cells was proven feasible as a more representative model of space in CA models. Results at the micro-scale level show that this new approach (there are very few examples of this in traditional CA model formulations, hence the classification of new) is as good as the one based on the use of regular cells as the ability of the CA models to simulate land use change is comparable with the existing, regular cell CA models. At the macro-scale, the use of irregular cells can be considered also feasible, as the macro-scale CA focus on aggregate measures derived from planning decisions.

The use of variable neighbourhoods also proved to be acceptable as a new modelling option when compared with the traditional user defined neighbourhoods in traditional CA models. The fact that the CA models at all scales are able to find a representative neighbourhood extent (in the form of a calibrated distance parameter), taking into consideration the coupled accessibility model (which is calibrated along with the neighbourhood parameters) makes them more representative of the spatial interactions (at both the land use and at the transport levels) than the ones that depend on a user knowledge to define the neighbourhood extent.

Calibration using the particle swarm (PS) algorithm also proven to be useful at dealing with the large number of calibration parameters that the presented CA models can sometimes have. When neighbourhood interaction rules are calibrated by the micro-scale CA model, by opposition to the more traditional approaches available in the literature where expert knowledge is in the base of this calibration, the complexity of the CA model increases, as more calibration parameters equal to more dimensions to the space of solutions, which becomes exponentially larger and more difficult to explore. The PS algorithm has proven to be very effective on dealing with this complexity, providing a simple conceptual process that also follows the simplicity of the CA models themselves. The PS implementation presented in this research works as a distributed processing procedure that is time consuming due to its brute force computing nature. Its temporal efficiency (an indicator commonly used in optimisation research) is therefore reduced, but its capacity of finding good sets of CA calibration parameters is noticeable. Algorithmic efficiency has to be measured in terms of the quality of the final results and not in terms of its processing efficiency in the case of land use CA models, as there is no time constraint in running the models (as it happens with, for example, traffic models).

The micro-scale CA model results compare well with existent CA models that aim to do land use allocation at local scales. Results from both the theoretical case studies and the real world case studies of Coimbra and those located on the Metropolitan Area of Barcelona present high levels of simulation accuracy, which are in line or are even higher than the comparable models reported in the literature. The use of a fitness measure that does not take into account land uses that cannot change is also an important improvement that highlights the quality of the micro-scale simulations.

The macro-scale CA model is innovative and has no comparable model in the CA literature (to the best of the candidate's knowledge at the time of the submission). There are other variable scale CA models that operate large cells but have a different rationale, transposing local transition rules to large scale cell structures, making it difficult to establish a direct comparison. The macro-scale CA model was able to achieve acceptable levels of accuracy maintaining the underling CA conceptual definition. The consideration of a dual neighbourhood definition proved to be feasible to model local competition for attracting population and jobs, representing both the existing tensions and cross benefits of thriving or deprived neighbouring municipalities.

Finally, the multi-scale CA model was developed to integrate the prospective functionalities of the two previous CA models. The integrated calibration of both scales was not feasible due to computational issues on the one hand, and the highly divergent nature of the capacity of achieving an acceptable level of accuracy, on the other hand. The very high number of variables, almost all of them structured in highly multi-dimensional arrays, was a very big constrain in terms of a possible calibration of a fully-fledged multi-scale CA model. On the other hand, lower levels of accuracy achieved especially by the macro-scale CA model, combined with the highly adequate but not perfect levels of accuracy of the micro-scale CA model, made the multi-scale CA model to diverge extremely from reality, providing very low and meaningless levels of accuracy. Therefore, the option was to just integrate the models on their prospective functionality, using calibration results from independent runs at both the micro- and the macro-scale levels. There is no loss of scientific validity and this issue is one of the topics for future research based on more efficient computational capacity and also on more detailed datasets that provide a finer grain resolution of retrospective historical data. The multi-scale CA model was nonetheless able to perform a valid prospective analysis, using just a small number of local scale case studies (again, due to lack of computational capacity to deal with larger datasets). The results show feasible future evolutions of land use change, illustrating the intended process of land use planning (at the macro-scale) coupled with land use allocation (at the micro-scale).

8.2 Meeting the research objectives

The research presented in this dissertation had set its aims at researching the use of CA models considering accessibility, considering the formulation of the CA mathematical concept as well as the maintenance of that formulation in all innovative proposals to the concept and, finally, building on the simplicity of the CA concept itself.

The inclusion of accessibility in all CA models developed with a fully integrated calibration between the land use processes and the accessibility component of the urban system proved to be feasible and successful in demonstrating that CA are also suitable as a modelling concept for the more common land use and transport interaction models. Although the way accessibility was represented in the models can be considered simple as just one mode was considered, this was mainly due to the computational complexity of including a multimodal accessibility model and can be easily tackled with more programming efficiency.

All the CA models presented are formulated according to the conceptual formulation of CA. The main CA components – cells, cell states, neighbourhood, transition rules and time – were modelled considering the requirements of CA, allowing the models to be fully compared with other CA models in the literature without any loss of conceptual robustness. The use of irregular cells complied with the definition of a cell and topological issues of neighbourhood and cell availability were maintained to follow the CA definition. The same happened with the definition of neighbourhood, with the introduction of variable neighbourhoods at all scales allowing the CA models to capture the extent of interaction at each scale as well as to include variable degrees of neighbourhood interaction considering different extents of the same neighbourhood as in the macro-scale CA model dual formulation. Transition rules were also kept under the formal CA formulation and were substantially improved with the use of the logit function to better classify cells in the land supply pool.

All these innovations were done under the strict observation of the conceptual formulation of the CA modelling concept, and kept the models simple and perceivable, a key feature of using CA as a spatial modelling tool, especially when dealing with land use and transport issues. This simplicity combined with the increase of representativeness brought by the different innovations on the CA components are the basis for the future development of models with greater potential of being accepted by practitioners, something that at this stage depends mainly on the computational efficiency of the modelling package.

As for the more detailed research objectives, the different components of the research achieved their main goals.

The comprehensive knowledge of the ongoing discussion on both modelling in urban studies and the use of CA as part of the toolkit to do it was the basis for the identification of the need to explore further the inclusion of accessibility as an endogenous component of the CA models as well as for the consideration of irregular cells and variable neighbourhoods as key innovations in spatial CA formulation. The development of the different components of both the micro-scale and the macro-scale CA models and their application to a set of different case studies demonstrate the potential of the use of CA in a multi-scale approach but also its main limitations, giving a contribution to the discussion about future modelling requirements and the way to implement these requirements considering the existing data management and computational capacities.

8.3 Future avenues of research

The results presented in this dissertation show evidence that the option for exploring further the issues of scale, cell form, neighbourhood and transitions rules have space in the candidate's future research agenda on CA. Also, the consideration of more complex formulations of accessibility as an endogenous component of the CA models is fully justified by the observed influence of this driver in land use change.

At the same time, there are important overarching issues that will be part of the candidate's future research agenda. Software development and implementation on the design side of the modelling packages, and model application in policy contexts on the outcome side are the two main areas of future interest.

The first overarching area of research will be the implementation of the CA modelling package in a more efficient, fully object-oriented programming language that allows a full deployment of the models regardless of problem size or model formulation, comprising as many modelling parameters as necessary under any type of calibration procedure. This new software implementation has to be designed in such an architecture that allows the use of multiple procedures and functions programmed in different languages, using common datasets and providing stand-alone mapping and visualisation techniques. This modelling environment must also be opened to the possibility of being fed by datasets coming from other models and of including other modelling functions written in script languages such as R (suitable to data processing) and Python (suitable for visualisation procedures). Nonetheless, this new implementation must keep the simplicity that characterises the current multi-scale CA model architecture, as well as a strong degree of universality, in order to be applied to multiple geographical and planning contexts. Efforts must be made to combine the expertise of computer scientists with the modelling expertise of the candidate, so to be focused on the modelling side and not on the computational side of the problems.

The second overarching issue regards the model application in real decision support contexts in planning and policy design. There is a small but vibrant set of recent scientific outputs in this field illustrating a need for robust demonstrations of the value of the use of urban modelling tools in planning, which is for a long time one of the key issues in urban planning as discussed in chapters 2 and 3. Future research will include comparative model applications to benchmark the multi-scale CA approach against the most established models in practice, as well as the development of a strategy to engage with practitioners in different planning and policy contexts to evaluate the robustness of the models.

On the CA modelling side all the main issues discussed in this dissertation remain valid as future research topics.

Scale is one of the key relevant topics in CA modelling. CA models have great potential in dealing with different spatial scales for modelling different phenomena, from the regional scale where spatial units are administrative or planning units that correspond to decision making units, to the very local neighbourhood scale where spatial units are the most disaggregated units of data, combining traditional data with new datasets coming from a wide variety of sensors, from traffic to consumption and energy. This varied set of scales will help to increase model representativeness and to engage with different practitioners and actors with different spatial interests.

Cell form is definitely an area of future research. The use of irregular cells has proven to be feasible and clearly provides better representation of spatial structures at every scale, allowing the consideration of both administrative and morphological spatial units. This finding is supported by the very recent trend on the development of CA models based on irregular cells, also named vector-based CA models, of which the work of Stevens and Dragicevic (2007), Moreno et al. (2008) and Moreno et al. (2009) and more recently Dahal and Chow (2015) are very good examples. Topological issues are in the centre of future research, opening a wider set of possibilities not only considering form but also the dynamics of the system (with evolving cell units throughout time) and neighbourhood interactions. The consideration of these dynamic irregular cells has many implications on the computational side of the models and need further research both in terms of the topological definition of cells and on the computational processing algorithms to deal with the much increased complexity of fully dealing with geometry within CA models.

Neighbourhood is the next topic to research. Neighbourhood definition is still mainly predetermined and the research reported in this dissertation showed the potential of using model-defined neighbourhoods rather than user-defined ones. Again, recent research trends can be found on the work of Crols et al. (2015) and Dahal and Chow (2015) illustrating the relevance of the findings of this research. Research will be conducted on neighbourhood definition, continuity and homogeneity, both on the topological side and also on the computational side. As for cells, more complex neighbourhood definitions increase substantially the computational complexity of some CA procedures increasing the necessity of more efficient algorithms.

The last CA component that will have the future focus of research is transition rules, which is illustrated by recent work on the topic (Wang et al., 2011, Yang et al., 2013, Li et al., 2014, Altartouri et al., 2015, Li et al., 2015). The multi-scale CA approach opens the possibility of designing different transition rules at different scales, thus enhancing the sensibility of CA to the particular spatial context modelled. Therefore, research is needed on the proper identification of drivers at all scales, as well as forms of parametrising these drivers. This implies the possibility of flexible testing of different phenomena with interactive modules for definition of transition rules without the need of background programming. There is also a need to investigate the robustness of transition rules by selecting subsets of the datasets (e.g. central urban areas, urban fringes, infrastructure corridors, all belonging to the same case study) to test if there is any variation on the calibration parameters, again working at separate scales to differentiate local from regional/metropolitan interactions.

Finally, calibration is a key component of the research in CA. The increasing appetite for using CA simplicity as a key advantage in the development of decision support systems for planning and decision making processes imply a stronger focus on the calibration of the models. The combination of the increased capacity to develop very capable simulation models with the incredible availability of data in the “big data” era highlights the need for better and more robust calibrations. The combination of other techniques and models, mainly coming from optimization sciences is here and there are many good recent examples of this coupling (Wang et al., 2011, Cao et al., 2013, Basse et al., 2014, Altartouri et al., 2015).

All these components of research have short- and mid- to long-term stages.

The most important short-term research must focus on the implementation of the current model architecture in a fully object-oriented structure, so new functionalities are available and larger datasets can be processed. This will contribute to the usability of the modelling package considering its use as a policy testing tool. Components of the multi-scale CA approach can be developed at this stage focusing on cell and neighbourhood topological formulation. The candidate has two research bids with these main aims currently being submitted to UK research councils.

The mid- to long-term research will have the new multi-scale CA modelling package as an environment for future development of several different story aspects of the CA formulation, based on both direct research and the future postgraduate research supervised by the candidate as a lecturer.

9 Bibliography

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P. & Schweik, C. M. (2002) A review and assessment of land-use change models: dynamics of space, time, and human choice. Newton Square, PA, U.S. Department of Agriculture, Forest Service, Northeastern Research Station.

Allen, E. (2001) INDEX: software for community indicators. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 229-262.

Almeida, C. M., Gleriani, J. M., Castejon, E. F. & Soares, B. S. (2008) Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22(9), 943-963.

Altartouri, A., Nurminen, L. & Jolma, A. (2015) Spatial neighborhood effect and scale issues in the calibration and validation of a dynamic model of Phragmites australis distribution – A cellular automata and machine learning approach. *Environmental Modelling & Software,* 71(9)**,** 15-29.

Barredo, J. I. & Demicheli, L. (2003) Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos. *Cities,* 20(5)**,** 297-310.

Barredo, J. I., Kasanko, M., Mccormick, N. & Lavalle, C. (2003) Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning,* 64(3)**,** 145-160.

Barros, J. (2005) Urban dynamics in Latin American cities: an agent-based simulation approach. *9th Conference on Computers in Urban Planning and Urban Management.* London, UK.

Basse, R. M., Omrani, H., Charif, O., Gerber, P. & Bódis, K. (2014) Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography,* 53(9)**,** 160-171.

Batty, M. (1994) A chronicle of scientific planning - The Anglo-American modelling experience. *Journal of the American Planning Association,* 60(1)**,** 7-16.

Batty, M. (2004) Dissecting the streams of planning history: technology versus policy trough models. *Environment and Planning B: Planning and Design,* 31(3)**,** 326-330.

Batty, M. (2005a) Agents, cells, and cities: new representational models for simulating multiscale urban dynamics. *Environment and Planning A,* 37(8)**,** 1373-1394.

Batty, M. (2005b) *Cities and complexity : understanding cities with cellular automata, agent-based models, and fractals.* Cambridge, MA, London: MIT Press.

Batty, M. (2013) *The New Science of Cities.* Cambridge, MA, London: MIT Press.

Batty, M., Couclelis, H. & Eichen, M. (1997) Editorial: Urban systems as cellular automata. *Environment and Planning B: Planning and Design,* 24(1)**,** 159-164.

Batty, M. & Torrens, P. (2001) Modelling complexity: the limits to prediction. *CASA Working Papers Series*.

Batty, M. & Xie, Y. (1994) From Cells to Cities. *Environment and Planning B-Planning & Design,* 21(Celebration Issue)**,** S31-S38.

Batty, M. & Xie, Y. (1997) Possible urban automata. *Environment and Planning B-Planning & Design,* 24(2)**,** 175-192.

Batty, M., Xie, Y. & Sun, Z. (1999) Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems,* 23(3)**,** 205-233.

Benenson, I. (2007) Warning! The scale of land-use CA is changing! *Computers Environment and Urban Systems,* 31(2)**,** 107-113.

Benenson, I., Aronovich, S. & Noam, S. (2005) Let's talk objects: generic methodology for urban high-resolution simulation. *Computers, Environment and Urban Systems,* 29(4)**,** 425-453.

Benenson, I., Omer, I. & Hatna, E. (2002) Entity-based modelling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design,* 29(4)**,** 491-512.

Benenson, I. & Torrens, P. M. (2004) *Geosimulation - Automata-based modeling of urban phenomena.* Chechester: John Wiley & Sons Ltd.

Benguigui, L., Czamanski, D., Marinov, M. & Portugali, Y. (2000) When and where is a city fractal? *Environment and Planning B: Planning and Design,* 27(4)**,** 507-519.

Briassoulis, H. (2008) Land-use policy and planning, theorizing, and modeling: lost in translation, found in complexity? *Environment and Planning B: Planning and Design,* 35(1)**,** 16-33.

Brömmelstroet, M. T. (2012) Transparency, flexibility, simplicity: From buzzwords to strategies for real PSS improvement. *Computers, Environment and Urban Systems,* 36(1)**,** 96-104.

Brömmelstroet, M. T. & Schrijnen, P. M. (2010) From planning support systems to mediated planning support: a structured dialogue to overcome the implementation gap. *Environment and Planning B-Planning & Design,* 37(1)**,** 3-20.

Candau, J. (2000) Calibrating a cellular automaton model of urban growth in a timely manner. *4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4).* Banff, Alberta, Canada.

Cao, K., Huang, B., Li, M. & Li, W. (2013) Calibrating a cellular automata model for understanding rural–urban land conversion: a Pareto front-based multi-objective optimization approach. *International Journal of Geographical Information Science,* 28(5)**,** 1028-1046.

Cao, M., Tang, G. A., Shen, Q. & Wang, Y. (2015) A new discovery of transition rules for cellular automata by using cuckoo search algorithm. *International Journal of Geographical Information Science,* 29(5)**,** 806-824.

Caruso, G., Peeters, D., Cavailhšs, J. & Rounsevell, M. (2009) Spaceÿ - ÿtime patterns of urban sprawl, a 1D cellular automata and microeconomic approach. *Environment and Planning B: Planning and Design,* 36(6)**,** 968-988.

Casti, J. (1997) *Would-be worlds: how simulation is changing the frontiers of science.* New York: John Wiley and Sons.

Clarke, K. (2002) Land Use Change Modeling Using SLEUTH. *Advanced Training Workshop on Land Use and Land Cover Change Study*. Taiwan, National Central University/National Taiwan University/ START.

Clarke, K., Hoppen, S. & Gaydos, L. (1996) Methods and Techniques for Rigourous Calibration of a Cellular Automaton Model of Urban Growth. *Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling*. Santa Fe, NM.

Clarke, K., Hoppen, S. & Gaydos, L. (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design,* 24(2)**,** 247-261.

Couclelis, H. (1985a) Cellular worlds - a framework for modeling micro-macro dynamics. *Environment and Planning A,* 17(5)**,** 585-596.

Couclelis, H. (1985b) Cellular worlds: a framework for modelling micro-macro dynamics. *Environment and Planning A,* 17585-596.

Couclelis, H. (1987) Cellular-dynamics - how individual decisions lead to global urban change. *European Journal of Operational Research,* 30(3)**,** 344-346.

Couclelis, H. (1997a) From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B: Planning and Design,* 24165-174.

Couclelis, H. (1997b) From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B-Planning & Design,* 24(2)**,** 165-174.

Couclelis, H. (2005) "Where has the future gone?" Rethinking the role of integrated land-use models in spatial planning. *Environment and Planning A,* 37(8)**,** 1353-1371.

Couto, P. (2003) Assessing the accuracy of spatial simulation models. *Ecological Modelling,* 167(1-2)**,** 181–198.

Crols, T., White, R., Uljee, I., Engelen, G., Poelmans, L. & Canters, F. (2015) A travel time-based variable grid approach for an activity-based cellular automata model. *International Journal of Geographical Information Science,* 29(10)**,** 1757-1781.

Dahal, K. R. & Chow, T. E. (2015) Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science,* 29(3)**,** 475-497.

De Keersmaecker, M.-L., Frankhauser, P. & Thomas, I. (2003) Using fractal dimensions for characterizing intra-urban diversity: the example of Brussels. *Geographical Analysis,* 35(4)**,** 310-328.

De La Barra, T. (2001) Integrated land use and transport modelling: the TRANUS experience. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 129-156.

Eloranta, K. (1997) Critical growth phenomena in cellular automata. *Physica D: Nonlinear Phenomena Lattice Dynamics,* 103(1-4)**,** 478-484.

Engelen, G., White, R., Uljee, I. & Drazan, P. (2002) The MURBANDY and MOLAND models for Dublin. Ispra, European Commission Joint Research Centre.

Feng, Y. & Liu, Y. (2012) A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing. *International Journal of Geographical Information Science,* 27(3)**,** 449-466.

Friedman, B., Kahn, P. & Borning, A. (2002) Value Sensitive Design: Theory and Methods. Seattle, University of Washington.

Gerling, R. W. (1990) Classification of triangular and honeycomb cellular automata. *Physica A: Statistical and Theoretical Physics,* 162(2)**,** 196-209.

He, C., Okada, N., Zhang, Q., Shi, P. & Li, J. (2008) Modelling dynamic urban expansion processes incorporating a potential model with cellular automata. *Landscape and Urban Planning,* 86(1)**,** 79-91.

Hunt, J. D. & Simmonds, D. C. (1993) Theory and Application of an Integrated Land-Use and Transport Modeling Framework. *Environment and Planning B-Planning & Design,* 20(2)**,** 221-244.

Ine (2001) *Censos 2001*. [online] Lisboa: Ine Available from www.ine.pt [Accessed n/a]

Kakaraparthi, S. K. & Kockelman, K. M. (2011) Application of UrbanSim to the Austin, Texas, Region: Integrated-Model Forecasts for the Year 2030. *Journal of Urban Planning and Development-Asce,* 137(3)**,** 238-247.

Kennedy, J. (1997) The particle swarm: Social adaptation of knowledge. In: (ed.^(eds.) *1997 IEEE International Conference on Evolutionary Computation (ICEC '97),* INDIANAPOLIS, IN 303-308.

Kennedy, J. & Eberhart, R. (1995) Particle swarm optimization. *Ieee International Conference on Neural Networks Proceedings, Vols 1-6,* 1942-1948.

Klosterman, R. (1994) Large-Scale urban models - Retrospect and prospect. *Journal of the American Planning Association,* 60(1)**,** 3-6.

Klosterman, R. & Petit, C. (2005) Editorial: An update on planning support systems. *Environment and Planning B: Planning and Design,* 32(4)**,** 477-484.

Klosterman, R. E. (2001) The What If? planning support system. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 263-284.

Klosterman, R. E., Siebert, L., Hoque, M. A., Kim, J. W. & Parveen, A. (2003) Using an operational planning support system to evaluate farmland preservation policies. In: GEERTMAN, S. & STILLWELL, J. (eds.) *Planning Support Systems in Practice.* 391-407.

Kocabas, V. & Dragicevic, S. (2006) Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems,* 30(6)**,** 921-953.

Koomen, E., Hilferink, M. & Borsboom-Van Beurden, J. (2011) Introducing Land Use Scanner. In: KOOMEN, E. & BORSBOOM-VAN BEURDEN, J. (eds.) *Land-Use Modelling in Planning Practice.* Springer Netherlands, 3-21.

Kwartler, M. & Bernar, R. N. (2001) CommunityViz: an integrated planning support system. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 285-308.

Landis, J. (2001) CUF, CUF II and CURBA: a family of spatially explicit urban growth and land-use policy simulation models. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 157-200.

Latuso, K. (2003) The SPARTACUS system for defining and analysing sustainable land use and transport policies. In: GEERTMAN, S. & STILLWELL, J. (eds.) *Planning Support Systems in Practice.* Heidelberg, Springer, 453-463.

Lavalle, C., Baranzelli, C., Silva, F. B. E., Mubareka, S., Gomes, C. R., Koomen, E. & Hilferink, M. (2011) A High Resolution Land Use/Cover Modelling Framework for

Europe: Introducing the EU-ClueScanner100 Model. In: MURGANTE, B., GERVASI, O., IGLESIAS, A., TANIAR, D. & APDUHAN, B. O. (eds.) *Computational Science and Its Applications - Iccsa 2011, Pt I.* Berlin, Springer-Verlag Berlin, 60-75.

Lee, D. B. (1973) Requiem for Large-Scale Models. *Journal of the American Institute of Planners,* 39(3)**,** 163-178.

Lee, D. B. (1994) Retrospective on Large-Scale Urban Models. *Journal of the American Planning Association,* 60(1)**,** 35-40.

Leontief, W. W. (1966) *Input-output economics.* New York,: Oxford University Press.

Levy, S. (1992) *Artificial life : the quest for a new creation.* New York: Pantheon Books.

Li, X., Lao, C., Liu, Y., Liu, X., Chen, Y., Li, S., Ai, B. & He, Z. (2013) Early warning of illegal development for protected areas by integrating cellular automata with neural networks. *Journal of Environmental Management,* 130(0)**,** 106-116.

Li, X. & Liu, X. (2006) An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region. *International Journal of Geographical Information Science,* 20(10)**,** 1109 - 1136.

Li, X., Liu, X. & Gong, P. (2015) Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model. *International Journal of Geographical Information Science,* 29(5)**,** 762-785.

Li, X., Liu, X. & Yu, L. (2014) A systematic sensitivity analysis of constrained cellular automata model for urban growth simulation based on different transition rules. *International Journal of Geographical Information Science,* 28(7)**,** 1317-1335.

Li, X. & Yeh, A. (2001a) Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environment and Planning A,* 331445-1462.

Li, X. & Yeh, A. G.-O. (2002) Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science,* 16(4)**,** 323 - 343.

Li, X. & Yeh, A. G. O. (2000) Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science,* 14(2)**,** 131-152.

Li, X. & Yeh, A. G. O. (2001b) Calibration of cellular automata by using neural networks for the simulation of complex urban systems. *Environment and Planning A,* 33(8)**,** 1445-1462.

Lindenmayer, A. (1968) Mathematical models for cellular interaction in development, Parts I and II. *Journal of Theoretical Biology,* 18(-)**,** 280-315.

Liu, X., Li, X., Shi, X., Wu, S. & Liu, T. (2008) Simulating complex urban development using kernel-based non-linear cellular automata. *Ecological Modelling,* 211(1-2)**,** 169-181.

Mcloughlin, J. B. (1969) *Urban and Regional Planning: A Systems Approach.* London: Faber and Faber.

Ménard, A. & Marceau, D. J. (2005) Exploration of spatial scale sensitivity in geographical cellular automata. *Environment and Planning B: Planning and Design,* 32(5)**,** 693-714.

Miller, E. J., Douglas Hunt, J., Abraham, J. E. & Salvini, P. A. (2004) Microsimulating urban systems. *Computers, Environment and Urban Systems,* 28(1–2)**,** 9-44.

Moreno, N., Ménard, A. & Marceau, D. J. (2008) VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design,* 35(8)**,** 647-665.

Moreno, N., Wang, F. & Marceau, D. J. (2009) Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems,* 33(1)**,** 44-54.

O' Sullivan, D. (2001a) Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis,* 33(1)**,** 1-18.

O' Sullivan, D. (2001b) *Graph-based cellular automaton models of urban spatial processes.* PhD. Bartlett School of Architecture and Planning, University College London.

O'sullivan, D. (2001) Exploring Spatial Process Dynamics Using Irregular Cellular Automaton Models. *Geographical Analysis,* 33(1)**,** 1-18.

Onsted, J. A. & Chowdhury, R. R. (2014) Does zoning matter? A comparative analysis of landscape change in Redland, Florida using cellular automata. *Landscape and Urban Planning,* 121(0)**,** 1-18.

Parsopoulos, K. E. & Vrahatis, M. N. (2002) Recent approaches to global optimization problems through Particle Swarm Optimization. *Natural Computing,* 1(2)**,** 235–306.

Petrov, L. O., Lavalle, C. & Kasanko, M. (2009) Urban land use scenarios for a tourist region in Europe: Applying the MOLAND model to Algarve, Portugal. *Landscape and Urban Planning,* 92(1)**,** 10-23.

Portugali, J., Meyer, H., Stolk, E. & Tan, E. (2012) *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design.* Springer Berlin Heidelberg.

Putman, S. H. & Shih-Liang, C. (2001) The METROPILUS planning support system: urban models and GIS. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 99-128.

Riks (2015) *Metronamica*. [online] Maastricht, The Netherlands: Riks Available from http://www.metronamica.nl/ [Accessed 2015/12/14]

Sakieh, Y., Salmanmahiny, A., Jafarnezhad, J., Mehri, A., Kamyab, H. & Galdavi, S. (2015) Evaluating the strategy of decentralized urban land-use planning in a developing region. *Land Use Policy,* 48534-551.

Santé, I., García, A. M., Miranda, D. & Crecente, R. (2010) Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning,* 96(2)**,** 108-122.

Semboloni, F. (1997) An urban and regional model based on cellular automata. *Environment and Planning B: Planning and Design,* 24(4)**,** 589-612.

Semboloni, F. (2000) The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B: Planning and Design,* 27(4)**,** 549-564.

Semboloni, F. (2005) Multi-agents simulation of urban dynamic. *9th Conference on Computers in Urban Planning and Urban Management.* London, UK.

Silva, E. & Clarke, K. C. (2002) Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems,* 26(6)**,** 525-552.

Silva, E. A. & Clarke, K. C. (2005) Complexity, emergence and cellular urban models: Lessons learned from applying sleuth to two Portuguese metropolitan areas. *European Planning Studies,* 13(1)**,** 93-116.

Silva, E. A., Healey, P., Harris, N. & Van Den Broeck, P. (2014) *The Routledge Handbook of Planning Research Methods.* Oxford, New York: Taylor & Francis.

Snyder, K. (2001) Tools for community design and decision-making. In: GEERTMAN, S. & STILLWELL, J. (eds.) *Planning Support Systems in Practice.* Hedelberg, Springer, 99-120.

Stevens, D. & Dragicevic, S. (2007) A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design,* 34(4)**,** 708-724.

Stevens, D., Dragicevic, S. & Rothley, K. (2007) iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software,* 22(6)**,** 761-773.

Straatemeier, T., Bertolini, L., Brömmelstroet, M. T. & Hoetjes, P. (2010) An experiential approach to research in planning. *Environment and Planning B-Planning & Design,* 37(4)**,** 578-591.

Swarm (2002) *Swarm - Multi-Agent Simulation Environment*. [online] Swarm Available from http://www.swarm.org/wiki/Main_Page [Accessed 2006/09/07]

Takeyama, M. & Couclelis, H. (1997) Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science,* 11(1)**,** 73-91.

Tobler, W. (1979) Cellular geography. In: GALE, S. & OLSSON, G. (eds.) *Philosophy in Geography.* Boston, D. Reidel, 379-386.

Tobler, W. R. (1970) Computer Movie Simulating Urban Growth in Detroit Region. *Economic Geography,* 46(2)**,** 234-240.

Torrens, P. (2000) How cellular models of urban systems work (1. Theory). In: BATTY, M. (ed.)) *CASA WP Series 28.* London, University College London, Torrens, P. & O' Sullivan, D. (2001) Editorial: Cellular automata and urban simulation: where do we go from here? *Environment and Planning B: Planning and Design,* 28163-168.

Van Den Bergh, F. & Engelbrecht, A. P. (2006) A study of particle swarm optimization particle trajectories. *Information Sciences,* 176(8)**,** 937-971.

Van Vliet, J., White, R. & Dragicevic, S. (2009) Modeling urban growth using a variable grid cellular automaton. *Computers Environment and Urban Systems,* 33(1)**,** 35-43.

Vancheri, A., Giordano, P., Andrey, D. & Albeverio, S. (2008) Urban growth processes joining cellular automata and multiagent systems. Part1: theory and models. *Environment and Planning B: Planning and Design,* 35(4)**,** 723-739.

Vandergue, D., Treuil, J.-P. & Drogoul, D. (2000) Modelling urban phenomena with cellular automata. In: BALLOT, G. & WEISBUCH, G. (eds.) *Applications of Simulation to Social Science.* Stanmore, Middlesex, Hermes Science Publishing, 127-140.

Verburg, P., Eickhout, B. & Van Meijl, H. (2008) A multi-scale, multi-model approach for analyzing the future dynamics of European land use. *The Annals of Regional Science,* 42(1)**,** 57-77.

Vito (2015) *Land Use Research Group*. [online] Mol, Belgium: Vito Available from https://vito.be/en/land-use [Accessed 2015/12/14]

Vliet, J. V., White, R. & Dragicevic, S. (2009) Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems,* 33(1)**,** 35-43.

Waddell, P. (2001) Between politics and planning: UrbanSim as a decision-support system for metropolitan planning. In: BRAIL, R. K. & KLOSTERMAN, R. E. (eds.) *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools.* Redlands, CA, ESRI Press, 201-228.

Waddell, P., Bhat, C., Ruiter, E., Bekhor, S., Outwater, M. & Schroer, E. L. (2001) Review of the Literature and Operational Models. Puget Sound Region, University of Washington.

Waddell, P. & Ulfarsson, G. (2004) Introduction to urban simulation: design and development of operational models. In: HENSHER, D., BUTTON, K., HAYNES, K. & STOPHER, P. (eds.) *Handbook of Transport Geography and Spatial Systems, Volume 5 (Handbooks in Transport).* 1st. Elsevier Science, 203-236.

Wagner, D. F. (1997) Cellular automata and geographic information systems. *Environment and Planning B: Planning and Design,* 24219-234.

Wang, F., Hasbani, J. G., Wang, X. & Marceau, D. J. (2011) Identifying dominant factors for the calibration of a land-use cellular automata model using Rough Set Theory. *Computers Environment and Urban Systems,* 35(2)**,** 116-125.

Ward, D., Murray, A. & Phinn, S. (2003) Integrating spatial optimization and cellular automata for evaluating urban change. *Annals of Regional Science,* 37(1)**,** 131-148.

Wegener, M. (1994) Operational urban models - State of the Art. *Journal of the American Planning Association,* 60(1)**,** 17-29.

White, R. (2007) Multi-scale modelling with variable grid CA. IN BAVAUD, F. & MAGER, C. (Eds.) *15th European Conference on Theoretical Quantitative Geography.* Montreux, Switzerland, Institute of Geography of the University of Lausanne.

White, R. & Engelen, G. (1993) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A,* 25(8)**,** 1175-1199.

White, R. & Engelen, G. (1997) Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B: Planning and Design,* 24(2)**,** 235-246.

White, R. & Engelen, G. (2000) High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems,* 24(5)**,** 383-400.

White, R., Engelen, G. & Uljee, I. (1997) The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B-Planning & Design,* 24(3)**,** 323-343.

White, R., Engelen, G. & Uljee, I. (2015) *Modeling Cities and regions as Complex Systems: From Theory to Planning Applications.* Cambridge, MA, London: MIT Press.

Wolfram, S. (1983) Statistical-Mechanics of Cellular Automata. *Reviews of Modern Physics,* 55(3)**,** 601-644.

Wolfram, S. (1984) Computation Theory of Cellular Automata. *Communications in Mathematical Physics,* 96(1)**,** 15-57.

Wu, F. (1998) SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science,* 12(1)**,** 63 - 82.

Wu, F. & Webster, C. J. (1998) Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design,* 25(1)**,** 103-126.

Xie, Y. C. (1996) A generalized model for cellular urban dynamics. *Geographical Analysis,* 28(4)**,** 350-373.

Yang, J., Tang, G. A., Cao, M. & Zhu, R. (2013) An intelligent method to discover transition rules for cellular automata using bee colony optimisation. *International Journal of Geographical Information Science,* 27(10)**,** 1849-1864.

Yang, Q. S., Li, X. & Shi, X. (2008) Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences,* 34(6)**,** 592-602.

Zeigler, B. P. (1976) *Theory of Modeling and Simulation.* New York: Wiley.